DISCUSSION PAPER SERIES

DP15305

EXCHANGE RATE PREDICTION WITH MACHINE LEARNING AND A SMART CARRY TRADE PORTFOLIO

Ilias Filippou, David Rapach, Mark Taylor and Guofu Zhou

INTERNATIONAL MACROECONOMICS AND FINANCE

ISSN 0265-8003

EXCHANGE RATE PREDICTION WITH MACHINE LEARNING AND A SMART CARRY TRADE PORTFOLIO

Ilias Filippou, David Rapach, Mark Taylor and Guofu Zhou

Discussion Paper DP15305
Published 17 September 2020
Submitted 15 September 2020

EXCHANGE RATE PREDICTION WITH MACHINE LEARNING AND A SMART CARRY TRADE PORTFOLIO

Abstract

We establish the out-of-sample predictability of monthly exchange rate changes via machine learning techniques based on 70 predictors capturing country characteristics, global variables, and their interactions. To guard against overfi tting, we use the elastic net to estimate a high-dimensional panel predictive regression and find that the resulting forecast consistently outperforms the naive no-change benchmark, which has proven difficult to beat in the literature. The forecast also markedly improves the performance of a carry trade portfolio, especially during and after the global financial crisis. When we allow for more complex deep learning models, nonlinearities do not appear substantial in the data.

JEL Classification: C45, F31, F37, G11, G12, G15

Keywords: exchange rate predictability, Elastic Net, carry trade, deep neural network

Ilias Filippou - iliasfilippou@wustl.edu
Washington University in St Louis

David Rapach - david.rapach@wustl.edu
Washington University in St Louis

Mark Taylor - mark.p.taylor@wustl.edu
Washington University in St Louis and CEPR

Guofu Zhou - zhou@wustl.edu
Washington University in St Louis

Exchange Rate Prediction with Machine Learning and a Smart Carry Trade Portfolio

Ilias Filippou
Washington University in St. Louis
iliasfilippou@wustl.edu

David E. Rapach
Washington University in St. Louis and Saint Louis University
david.rapach@wustl.edu

Mark P. Taylor
Washington University in St. Louis and Centre for Economic Policy Research
mark.p.taylor@wustl.edu

Guofu Zhou*
Washington University in St. Louis and China Academy of Financial Research
zhou@wustl.edu

September 1, 2020

*Corresponding author. Send correspondence to Guofu Zhou, Olin Business School, Washington University in St. Louis, St. Louis, MO 63130; e-mail: zhou@wustl.edu; phone: 314-935-6384. We thank conference and seminar participants at the 2019 International Symposium on Forecasting, NIESR/CFM/OMFIF 2019 Workshop on Modeling the Macroeconomy in Risky Times, 2019 City University of Hong Kong Workshop in Econometrics and Statistics, Aarhus University, University of North Carolina at Charlotte, and Washington University in St. Louis, as well as Yufeng Han, Ai He, Dashan Huang, and Jiahan Li, for valuable comments.

Exchange Rate Prediction with Machine Learning and a Smart Carry Trade Portfolio

Abstract

We establish the out-of-sample predictability of monthly exchange rate changes via machine learning techniques based on 70 predictors capturing country characteristics, global variables, and their interactions. To guard against overfitting, we use the elastic net to estimate a high-dimensional panel predictive regression and find that the resulting forecast consistently outperforms the naïve no-change benchmark, which has proven difficult to beat in the literature. The forecast also markedly improves the performance of a carry trade portfolio, especially during and after the global financial crisis. When we allow for more complex deep learning models, nonlinearities do not appear substantial in the data.

JEL classifications: C45, F31, F37, G11, G12, G15

Key words: Short-horizon exchange rate predictability, Panel predictive regression, Elastic net, Carry trade, Deep neural network

1 Introduction

The specter of Meese and Rogoff (1983) continues to haunt international finance: despite an array of theoretical models linking fundamentals to exchange rates, it is difficult to consistently outperform the naïve random walk without drift (or no-change) exchange rate forecast on an out-of-sample basis, especially at short horizons. Of course, it is not surprising that it is difficult to predict exchange rate changes. Developed-country currencies are traded in quite liquid and institutional investor-dominated markets, which are reasonably efficient; we thus expect that exchange rate fluctuations will contain at most only a small predictable component. The same, however, is true for equities, and while we also expect a small predictable component in equity returns, the apparent consensus in the literature is that short-horizon, out-of-sample stock return predictability exists to a statistically and economically significant degree (Rapach and Zhou 2013). Such a consensus does not prevail with respect to exchange rate changes, as the empirical evidence for short-horizon, out-of-sample exchange rate predictability appears considerably weaker and more precarious (Rossi 2013).

In this paper, we investigate the ability of machine learning techniques to improve monthly out-of-sample forecasts of US dollar exchange rates for 14 developed countries.[1] To predict exchange rates, we use a rich information set comprised of ten country characteristics and six global variables. The country characteristics capture a diversity of information that is potentially relevant to foreign exchange market participants. The characteristics include various macroeconomic and financial variables, such as inflation, unemployment gap, interest rate, and financial valuation ratio differentials, which can be motivated by the Taylor (1993) rule, uncovered interest parity (UIP), and uncovered equity parity (UEP), among other theories. The global variables include economic and monetary policy uncertainty indices (Baker, Bloom, and Davis 2016), a geopolitical risk index (Caldara and Iacoviello 2018), as well as

[1] Machine learning is becoming popular in finance for analyzing equity returns with large information sets (e.g., Rapach, Strauss, and Zhou 2013; Feng, Giglio, and Xiu 2020; Freyberger, Neuhierl, and Weber 2020; Gu, Kelly, and Xiu 2020; Han et al. 2020; Kozak, Nagel, and Santosh 2020).

measures of global foreign exchange volatility, illiquidity, and correlation (Menkhoff et al. 2012a; Mueller, Stathopoulos, and Vedolin 2017). To allow for the predictive relationships between the country characteristics and exchange rate changes to vary with global conditions, we interact the country characteristics with the global variables, producing a set of 70 predictors. The predictors serve as explanatory variables in a predictive regression for (proportional) exchange rate changes.

Given our high-dimensional setting with 70 predictors, it is vital to guard against *overfitting* when estimating the predictive regression. The need to alleviate overfitting is made more urgent by the inherently large unpredictable component in monthly exchange rate changes, which means that we need to contend with noisy data when estimating predictive relationships. To guard against overfitting, we estimate the predictive regression via the *elastic* net (ENet, Zou and Hastie 2005), a refinement of the popular *least absolute shrinkage and selection operator* (LASSO, Tibshirani 1996) from machine learning. By construction, conventional ordinary least squares (OLS) estimation maximizes the fit of the model over the estimation (or training) sample, which can lead to overfitting the model to the training sample and thus poor out-of-sample performance, especially for high-dimensional models and noisy data. The ENet is a *penalized regression* technique that shrinks the estimated coefficients toward zero, thereby alleviating overfitting. The penalty term in the ENet includes both an ℓ_1 component—as in the LASSO—and an ℓ_2 component—as in ridge regression (Hoerl and Kennard 1970); the former permits shrinkage to zero, so that the ENet also performs variable selection. We further guard against overfitting by estimating the predictive regression in a panel framework, in which we pool the data. Pooling substantially reduces the number of parameters that we need to estimate, thereby helping to improve out-of-sample performance in light of the bias-variance tradeoff.

In implementing the ENet, a key issue is selecting (or tuning) the regularization parameter, which controls the degree of shrinkage. Although K-fold cross validation is the most popular method for tuning the regularization parameter, the number and construction of the folds are largely arbitrary. To better ensure a sufficient degree of shrinkage, we tune the regularization parameter using variants of the Bayesian information criterion (BIC, Schwarz 1978), including those developed by Wang, Li, and Leng (2009), Fan and Tang (2013), and Hui, Warton, and Foster (2015) for machine learning.

Simulating the situation of a forecaster in real time, we generate monthly out-of-sample forecasts based on the 70 predictors by recursively estimating the panel predictive regression via the ENet. Based on data availability and after allowing for a ten-year initial training sample, the out-of-sample period spans 1995:01 to 2019:03. We compute forecasts for the entire out-of-sample period for the United Kingdom, Switzerland, Japan, Canada, Australia, New Zealand, Sweden, Norway, and Denmark; for the Euro area, the out-of-sample period begins in 2000:02. We refer to the exchange rates for these ten countries as the G10.[2] For Germany, Italy, France, and the Netherlands, the out-of-sample period ends in 1998:12, corresponding with their adoption of the Euro.

We find that the ENet exchange rate forecasts generally outperform the naïve no-change benchmark forecast over the 1995:01 to 2019:03 out-of-sample period. The ENet forecast that tunes the regularization parameter via the extended regularization information criterion (ERIC; Hui, Warton, and Foster 2015)—and which imposes the strongest degree of shrinkage—performs the best overall, demonstrating the importance of guarding against overfitting. The ENet-ERIC forecast outperforms the no-change benchmark in terms of mean squared forecast error (MSFE) for 13 of the 14 countries, including all of the G10 countries.[3] According to the Clark and West (2007) statistic, the improvement in MSFE is significant for most of the countries. Based on the Campbell and Thompson (2008) out-of-sample R^2 statistic, the improvements in forecast accuracy are also quantitatively large in the context of the extensive literature surveyed by Rossi (2013), with the monthly out-

[2]The G11 currencies are the US dollar, Euro, British pound, Swiss franc, Japanese yen, Canadian dollar, Australian dollar, New Zealand dollar, Swedish krona, Norwegian krone, and Danish krone. With the US dollar serving as the base currency, we label our set of ten exchange rates the G10.

[3]The conventional OLS forecast is plagued by overfitting and substantially underperforms the no-change benchmark in terms of MSFE.

of-sample R^2 statistic reaching as high as 5.32% for the United Kingdom. For the entire group of 14 countries taken together, the out-of-sample R^2 statistic is 2.04%, which constitutes a significant improvement in MSFE vis-à-vis the no-change benchmark according to the Clark and West (2007) test (at the 1% level). Based on the graphical device of Goyal and Welch (2003, 2008), the ENet-ERIC forecast outperforms the no-change benchmark on a reasonably consistent basis over time. The outperformance is especially strong during the worst phase of the global financial crisis in late 2008. Overall, by utilizing a rich information set—while adequately guarding against overfitting—the ENet-ERIC approach makes considerable progress in solving the Meese and Rogoff (1983) no-predictability puzzle and provides among the best short-horizon exchange rate forecasts available to date.

By performing variable selection, the ENet is a machine learning tool that facilitates model interpretation. We examine the recursive ENet coefficient estimates to identify the most relevant predictors of US dollar exchange rates for our group of developed countries. Among the individual country characteristics, financial valuation ratio differentials (especially the dividend yield differential) are among the most frequently selected predictors. The ENet also frequently selects a number of predictors involving interactions between the country characteristics and global variables. In particular, the interactions of inflation, the unemployment gap, and government bill and bond yield differentials with global foreign exchange volatility are relevant predictors according to the ENet. When we examine the recursive coefficient estimates, an interesting pattern emerges: the predictive relationships implied by theory (e.g., UIP) often become stronger as global foreign exchange volatility increases.

A spate of recent papers (e.g., Engel and Wu 2019; Jiang, Krishnamurthy, and Lustig 2019; Kremens and Martin 2019; Adrian and Xie 2020; Lilley et al. forthcoming) finds evidence of exchange rate predictability around the global financial crisis using variables related to the US dollar's "safe-haven" status. Based on data availability, these studies use relatively short samples and/or analyze medium- to long-horizon predictability (horizons of one quarter to multiple years). In contrast, we consider a lengthy out-of-sample period (beginning well before the global financial crisis) and focus on short-horizon (monthly) predictability, which has proven even more difficult to uncover in the literature (e.g., Mark 1995; Rossi 2013). Nevertheless, our results appear to capture exchange rate predictability during the crisis corresponding to the US dollar's safe-haven role. First, as mentioned previously, our monthly ENet forecasts generate strong out-of-sample gains during the crisis, consistent with the US dollar's safe-haven status contributing to increased exchange rate predictability around the crisis as global risk tolerance declined. Second, we find that, as a group, the predictors selected by the ENet are significantly related to the change in US holdings of foreign bonds, which Lilley et al. (forthcoming) use to measure capital flows. They find that their capital flow measure (which is available at the quarterly frequency) reflects global risk appetite and is significantly related to the change in a broad US dollar exchange rate index after 2006. After aggregating our monthly predictors over time to the quarterly frequency and across countries, we find that the most relevant predictors according to the ENet are significantly related to the Lilley et al. (forthcoming) capital flow measure. In sum, our evidence of monthly out-of-sample exchange rate predictability for the 1995:01 to 2019:03 period appears to capture the predictability stemming from the US dollar's safe-haven role around the crisis.

We also explore the implications of out-of-sample exchange rate predictability for carry trade investment strategies. The popular carry trade entails going long (short) currencies with relatively high (low) interest rates. Consider a US investor who goes long (short) the currency for country i (US dollar). Based on covered interest parity (CIP), we can approximate the excess return for the investment as

$$RX_{i,t+1} \approx (r_{i,t} - r_{\mathrm{US},t}) - \delta_{i,t+1}, \tag{1.1}$$

where $r_{i,t}$ ($r_{\text{US},t}$) is the month-t government bill yield for country i (the United States), $\delta_{i,t+1} = (S_{i,t+1}/S_{i,t}) - 1$ is the (proportional) change in the exchange rate, and $S_{i,t}$ is the spot exchange rate expressed as the number of country-i currency units per US dollar. Beginning with Hansen and Hodrick (1980), Bilson (1981), and Fama (1984), a voluminous literature finds that UIP does not hold and that the conditional expectation of $RX_{i,t+1}$ is positive in Equation (1.1) when $r_{i,t} - r_{\text{US},t} > 0$.[4] Compared to a variety of investment strategies, conventional carry trade portfolios deliver impressive Sharpe ratios (e.g., Burnside, Eichenbaum, and Rebelo 2011; Lustig, Roussanov, and Verdelhan 2011).[5] However, carry trade strategies suffered large losses in late 2008 , and their performance has deteriorated in the wake of the global financial crisis (e.g., Melvin and Taylor 2009; Jordà and Taylor 2012; Daniel, Hodrick, and Lu 2017; Melvin and Shand 2017).[6]

From the perspective of Equation (1.1), if an investor's forecast of $\delta_{i,t+1}$ is zero—in the spirit of Meese and Rogoff (1983)—then the investor's forecast of $RX_{i,t+1}$ is simply the bill yield differential ($r_{i,t} - r_{\text{US},t}$)—in the spirit of the carry trade. We first construct an optimal portfolio for a mean-variance investor who allocates across the 14 foreign currencies in our sample by relying on the bill yield differential to forecast the foreign currency excess return. We label this a *Basic Optimal* (Basic-Opt) carry trade portfolio, as the investor ignores exchange rate predictability when forecasting the excess return. The Basic-Opt portfolio delivers impressive performance before the global financial crisis. However, it suffers large losses in late 2008, and its cumulative return is essentially flat thereafter.[7]

We then construct a *Smart Optimal* (Smart-Opt) carry trade portfolio, where the mean-variance investor augments the bill yield differential with the ENet-ERIC forecast of the

[4]See Froot and Thaler (1990), Taylor (1995), and Burnside (2018) for surveys of UIP.

[5]Studies that explore risk-based explanations for carry trade returns include Burnside et al. (2011), Lustig, Roussanov, and Verdelhan (2011), Menkhoff et al. (2012a), Dobrynskaya (2014), Jurek (2014), Lettau, Maggiori, and Weber (2014), and Dahlquist and Hasseltoft (2020).

[6]Brunnermeier, Nagel, and Pedersen (2009) provide a explanation for carry trade crashes based on funding-constrained speculators.

[7]Daniel, Hodrick, and Lu (2017) also construct an optimal carry trade portfolio for a mean-variance investor who uses the interest rate differential to forecast the currency excess return; our Basic-Opt portfolio performs similarly to theirs. A conventional carry trade portfolio that sorts currencies based on the bill yield differential and goes long (short) the fifth (first) quintile performs even worse than the Basic-Opt portfolio.

exchange rate change to forecast the currency excess return.[8] Like the Basic-Opt portfolio, the Smart-Opt portfolio delivers impressive performance before the crisis. In sharp contrast to the Basic-Opt portfolio, it also performs impressively thereafter. Specifically, the Smart-Opt portfolio experiences a smaller loss in September of 2008, generates large gains in the last three months of 2008, and performs well subsequently. Consistent with a loss of risk appetite during the crisis and the US dollar's safe-haven role, the ENet forecast predicts a substantial depreciation for many foreign currencies in late 2008, which leads to markedly different allocations for the Smart-Opt vis-à-vis the Basic-Opt portfolio. The Smart-Opt portfolio also generates substantial alpha before and after the crisis in the context of the Lustig, Roussanov, and Verdelhan (2011) currency factor model.

Finally, we use deep neural networks (DNNs) to explore the relevance of complex nonlinear relationships for out-of-sample exchange rate prediction.[9] Our baseline panel predictive regression allows for a type of nonlinearity via the interactions of the country characteristics with the global variables; the model, however, remains linear in the predictors. DNNs are popular machine learning tools that allow for complex nonlinear predictive relationships via a network architecture with multiple hidden layers containing neurons activated by predictive signals. In order to approximate general predictive relationships, DNNs are highly parameterized, which makes them susceptible to overfitting, and the hyperparameters for the algorithms used to estimate DNNs can be set to better guard against overfitting. We investigate the performance of DNNs for forecasting exchange rates based on our set of 70 predictors. Although the DNN forecasts outperform the no-change benchmark for most countries, they do not generally perform as well as the ENet-ERIC forecast. Using the variable importance measure (Greenwell, Boehmke, and McCarthy 2018) and partial dependence plot (Friedman 2001) to peer inside the "black box" of the fitted DNNs, we find that the

[8] The results are similar when the investor uses the other ENet forecasts. Della Corte, Sarno, and Tsiakas (2009) construct mean-variance optimal portfolios for a US investor who allocates across US, British, German, and Japanese bonds using a handful of fundamentals to forecast exchange rates. Jordà and Taylor (2012) use a small number of fundamentals to improve carry strategies (but not in a mean-variance optimal framework).

[9] Gu, Kelly, and Xiu (2020) recently find some success in using DNNs to forecast monthly stock returns based on a large number of firm characteristics.

nonlinearities are relatively weak. When it comes to exchange rate forecasting, nonlinearities thus do not appear strong enough to offset the risk of overfitting associated with estimating a highly parameterized DNN in a noisy data environment.

The remainder of the paper is organized as follows. Section 2 describes the data. Section 3 discusses the specification and estimation of the panel predictive regression used to generate the out-of-sample exchange rate forecasts. Section 4 provides results for forecast accuracy. Section 5 discusses the construction of the Basic-Opt and Smart-Opt carry trade portfolios and analyzes their performance. Section 6 focuses on the DNN forecasts. Section 7 concludes.

2 Data

This section describes the data used in our analysis. Section A1 of the Internet Appendix provides further information on the construction of the variables and data sources.

2.1 Exchange Rates

We begin with daily exchange rate data from Barclays and Reuters via Datastream, and we convert daily spot exchange rates to a monthly frequency using end-of-month values (e.g., Burnside et al. 2011; Lustig, Roussanov, and Verdelhan 2011). Our sample consists of the following 14 countries: the United Kingdom, Switzerland, Japan, Canada, Australia, New Zealand, Sweden, Norway, Denmark, the Euro area, Germany, Italy, France, and the Netherlands.[10] Germany, Italy, France, and the Netherlands are replaced by the Euro area after the Euro's introduction in 1999. We refer to the group of ten countries excluding Germany, Italy, France, and the Netherlands as the G10.

As in Section 1, we use $S_{i,t}$ to denote the month-t spot exchange rate, expressed as the number of country-i currency units per US dollar (e.g., Lustig, Roussanov, and Verdelhan

[10] Our universe of exchange rates is similar to the sample of developed countries employed in other studies (e.g., Lustig, Roussanov, and Verdelhan 2011; Menkhoff et al. 2012a), with the exception of Belgium, which we exclude due to a lack of data availability for the country characteristics in Section 2.2.

2011; Menkhoff et al. 2012a,b). An increase in $S_{i,t}$ thus represents an appreciation of the US dollar. The country-i (proportional) change in the exchange rate is given by $\delta_{i,t} = (S_{i,t}/S_{i,t-1}) - 1$.

Table 1 reports summary statistics for the 14 exchange rate changes. The second column reports the sample period for each country. With the exceptions of four countries, the sample ends in 2019:03; for France, Germany, Italy, and the Netherlands, the sample ends in 1998:12, the last month for which these countries had their own currencies. Based on data availability for the predictors, the sample begins in 1985:01 for all countries, with the exception of the Euro area, where the sample begins in 1999:02, corresponding to the introduction of the Euro in January of 1999.

The annualized means in the third column of Table 1 are generally small in magnitude. In fact, none of the means is significant at conventional levels, so that, following many studies, we treat the average exchange rate change as zero for each country in our empirical modeling. The annualized volatilities in the fourth column are typically sizable; apart from Canada (7.39%), they range from 9.80% (Euro area) to 12.18% (New Zealand). The autocorrelations in the last column are all relatively small in magnitude. Overall, the summary statistics in Table 1 reflect well-known empirical features of exchange rates.

2.2 Country Characteristics

We consider ten monthly country characteristics computed using macroeconomic and financial data from Global Financial Data and the Organization for Economic Cooperation and Development:

Inflation differential ($\mathrm{INF}_{i,t}$) Difference in consumer price index (CPI) inflation rates for country i and the United States.

Unemployment gap differential ($\mathrm{UN}_{i,t}$) Difference in unemployment gaps for country i and the United States. The unemployment gap is the cyclical component of the unemployment rate computed using the Christiano and Fitzgerald (2003) band-pass filter for periodicities between six and 96 months.[11]

Bill yield differential ($\text{BILL}_{i,t}$) Difference in three-month government bill yields for country i and the United States.

Note yield differential ($\text{NOTE}_{i,t}$) Difference in five-year government note yields for country i and the United States.

Bond yield differential ($\text{BOND}_{i,t}$) Difference in ten-year government bond yields for country i and the United States.

Dividend yield differential ($\text{DP}_{i,t}$) Difference in dividend yields for country i and the United States.

Price-earnings differential ($\text{PE}_{i,t}$) Difference in price-earnings ratios for country i and the United States.

Stock market momentum ($\text{SRET}_{i,t}$) Difference in the cumulative twelve-month stock market return for country i and the United States.

Idiosyncratic volatility ($\text{IV}_{i,t}$) Integrated volatility computed for the fitted residuals from the Lustig, Roussanov, and Verdelhan (2011) two-factor model estimated using daily data for month t for country-i currency excess returns.

Idiosyncratic skewness ($\text{IS}_{i,t}$) Integrated skewness computed for the fitted residuals from the Lustig, Roussanov, and Verdelhan (2011) two-factor model estimated using daily data for month t for country-i currency excess returns.

[11] We compute the unemployment gap only using data available at the time of forecast formation. The Hodrick and Prescott (1997) filter is often used to compute output and unemployment gaps; however, it tends to perform poorly at the right-hand endpoint. Because we are interested in out-of-sample forecasting, we use the Christiano and Fitzgerald (2003) band-pass filter, which performs better at the right-hand endpoint.

The country characteristics include a variety of macroeconomic and financial measures, all of which are based on data readily available to foreign exchange market participants.[12] The inflation and unemployment gap differentials constitute Taylor (1993) rule fundamentals (e.g., Engel and West 2005; Molodtsova and Papell 2009), which appear to perform better for exchange rate prediction than fundamentals based on the traditional monetary model (Frenkel 1976; Mussa 1976). The bill yield differential relates to the voluminous UIP literature, while longer-term yield differentials are considered by Ang and Chen (2011) and Chen and Tsang (2013) in the context of yield curves. Hau and Rey (2006) and Cenedese et al. (2016), among others, employ valuation ratio differentials to analyze UEP. The other country characteristics represent additional financial measures that are potentially relevant to market participants.

2.3 Global Variables

We consider six global variables:

Economic policy uncertainty (EPU_t) Baker, Bloom, and Davis (2016) economic policy uncertainty index based on coverage frequencies in ten major US newspapers.

Monetary policy uncertainty (MPU_t) Baker, Bloom, and Davis (2016) monetary policy uncertainty index based on coverage frequencies in ten major US newspapers.

Geopolitical risk (GR_t) Caldara and Iacoviello (2018) geopolitical risk index based on newspaper coverage.

Global foreign exchange volatility (GVOL_t) Following Menkhoff et al. (2012a), global foreign exchange volatility is the average for the month of the daily cross-sectional averages of the absolute values of currency returns.

[12] We account for the publication lag in the CPI and unemployment rate, so that $\mathrm{INF}_{i,t}$ and $\mathrm{UN}_{i,t}$ correspond to data for month $t-1$ that are reported in month t.

Global foreign exchange illiquidity (GILL_t) Following Menkhoff et al. (2012a), global foreign exchange illiquidity is the average for the month of the daily cross-sectional averages of the bid-ask spreads for currency returns.

Global foreign exchange correlation (GCOR_t) Similarly to Mueller, Stathopoulos, and Vedolin (2017), we measure global foreign exchange correlation as the average of the realized covariances for all currency pairs computed using daily returns for the month.

The global variables capture general economic conditions that potentially affect the predictive ability of the country characteristics.[13]

3 Panel Predictive Regression

This section specifies the panel predictive regression and describes the construction of the out-of-sample forecasts.

3.1 Specification

We collect the month-t characteristics for country i and the month-t global variables in the following vectors:

$$\underset{(Z\times 1)}{\boldsymbol{z}_{i,t}} = [\ \text{INF}_{i,t}\quad \text{UN}_{i,t}\quad \text{BILL}_{i,t}\quad \text{NOTE}_{i,t}\quad \text{BOND}_{i,t}\quad \text{DP}_{i,t}\quad \text{PE}_{i,t}\quad \text{SRET}_{i,t}\quad \text{IV}_{i,t}\quad \text{IS}_{i,t}\ ]', \tag{3.1}$$

$$\underset{(G\times 1)}{\boldsymbol{g}_{t}} = [\ \text{EPU}_t\quad \text{MPU}_t\quad \text{GR}_t\quad \text{GVOL}_t\quad \text{GILL}_t\quad \text{GCOR}_t\ ]', \tag{3.2}$$

respectively, for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where N (T) is the number of countries (time-series observations). The vector of predictors for country i is comprised of the country

[13] We follow Menkhoff et al. (2012a) and Mueller, Stathopoulos, and Vedolin (2017) by measuring GVOL_t, GILL_t, and GCOR_t as the residuals from fitted first-order autoregressive processes. We only use data available at the time of forecast formation when fitting the autoregressive processes and computing the residuals. Bakshi and Panayotov (2013) and Filippou and Taylor (2017) use aggregated country characteristics and global variables to predict conventional carry trade portfolio returns, while we forecast individual exchange rate changes.

characteristics and the characteristics interacted with each global variable:

$$\underset{(K\times 1)}{\boldsymbol{x}_{i,t}} = [\ \boldsymbol{z}_{i,t}' \quad \boldsymbol{h}_{i,t}'\ ]', \tag{3.3}$$

where

$$\underset{(ZG\times 1)}{\boldsymbol{h}_{i,t}} = \boldsymbol{z}_{i,t} \otimes \boldsymbol{g}_t, \tag{3.4}$$

$\otimes$ is the Kronecker product, and $K = Z(G+1)$. Since $Z = 10$ and $G = 6$ in Equations (3.1) and (3.2), respectively, we have $K = 70$ predictors for each country. We express the country-i predictors in deviations from country-specific means:

$$\tilde{\boldsymbol{x}}_{i,t} = \boldsymbol{x}_{i,t} - \bar{\boldsymbol{x}}_{i\cdot}, \tag{3.5}$$

where $\bar{\boldsymbol{x}}_{i\cdot} = (1/T)\sum_{t=1}^{T} \boldsymbol{x}_{i,t}$. The panel predictive regression is given by

$$\delta_{i,t} = \tilde{\boldsymbol{x}}_{i,t-1}\boldsymbol{b} + \varepsilon_{i,t} \quad \text{for } i = 1,\ldots,N;\ t = 1,\ldots,T, \tag{3.6}$$

where $\boldsymbol{b} = [\ b_1 \quad \cdots \quad b_K\ ]'$ is the K-vector of slope coefficients and $\varepsilon_{i,t}$ is a zero-mean disturbance term. Because we set the country-specific mean for $\delta_{i,t}$ to zero, Equation (3.6) is tantamount to a fixed-effects specification.[14]

It is convenient to express the panel predictive regression in matrix notation as

$$\boldsymbol{\delta} = \tilde{\boldsymbol{X}}\boldsymbol{b} + \boldsymbol{\varepsilon}, \tag{3.7}$$

where

$$\underset{(NT\times 1)}{\boldsymbol{\delta}} = [\ \boldsymbol{\delta}_{1\cdot}' \quad \cdots \quad \boldsymbol{\delta}_{N\cdot}'\ ]', \tag{3.8}$$

[14] Hjalmarsson (2010) predicts global stock returns using a low-dimensional panel predictive regression, while Gu, Kelly, and Xiu (2020) forecast individual stock returns using a high-dimensional panel predictive regression estimated using a variety of machine learning techniques.

$$\underset{(T\times 1)}{\boldsymbol{\delta}_{i\cdot}} = [\ \delta_{i,1} \quad \cdots \quad \delta_{i,T}\ ]', \tag{3.9}$$

$$\underset{(NT\times K)}{\tilde{\boldsymbol{X}}} = [\ \tilde{\boldsymbol{X}}_{1\cdot}' \quad \cdots \quad \tilde{\boldsymbol{X}}_{N\cdot}'\ ]', \tag{3.10}$$

$$\underset{(T\times K)}{\tilde{\boldsymbol{X}}_{i\cdot}} = [\ \tilde{\boldsymbol{x}}_{i,0} \quad \cdots \quad \tilde{\boldsymbol{x}}_{i,T-1}\ ]', \tag{3.11}$$

$$\underset{(NT\times 1)}{\boldsymbol{\varepsilon}} = [\ \boldsymbol{\varepsilon}_{1\cdot}' \quad \cdots \quad \boldsymbol{\varepsilon}_{N\cdot}'\ ]', \tag{3.12}$$

$$\underset{(T\times 1)}{\boldsymbol{\varepsilon}_{i\cdot}} = [\ \varepsilon_{i,1} \quad \cdots \quad \varepsilon_{i,T}\ ]'. \tag{3.13}$$

For simplicity, we assume a balanced panel in the notation. When we estimate $\boldsymbol{b}$ in Equation (3.7) using our data, we have an unbalanced panel; it is straightforward to adjust the notation accordingly.

We include many more predictors than existing studies of exchange rate predictability, which typically consider a only limited number of country characteristics. In addition to numerous country characteristics, we include interactions of the country characteristics with a set of global variables.

3.2 Forecast Construction

An out-of-sample forecast of $\delta_{i,T+1}$ based on the panel predictive regression in Equation (3.6) and data available through T is given by

$$\hat{\delta}_{i,T+1|T}^{\mathrm{PPR}} = \tilde{\boldsymbol{x}}_{i,T}' \hat{\boldsymbol{b}}_{1:T}, \tag{3.14}$$

where $\hat{\boldsymbol{b}}_{1:T}$ is an estimate of $\boldsymbol{b}$ in Equation (3.6) based on data through T. Because the predictable component in monthly exchange rate changes is inherently small and the panel predictive regression is high dimensional, conventional OLS estimation of $\boldsymbol{b}$ is susceptible to overfitting.

The LASSO (Tibshirani 1996) is a popular machine learning technique based on penalized (or regularized) regression. It mitigates overfitting by including an ℓ_1 penalty term in the objective function for estimating $\boldsymbol{b}$ in Equation (3.7):

$$\underset{\boldsymbol{b}\in\mathbb{R}^K}{\arg\min}\ \frac{1}{2NT}\|\boldsymbol{\delta} - \tilde{\boldsymbol{X}}\boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{b}\|_1, \tag{3.15}$$

where $\lambda \geq 0$ is a regularization parameter that governs the degree of shrinkage and

$$\|\boldsymbol{v}\|_1 = \sum_{j=1}^{J}|v_j|, \tag{3.16}$$

$$\|\boldsymbol{v}\|_2 = \left(\sum_{j=1}^{J} v_j^2\right)^{0.5} \tag{3.17}$$

are the ℓ_1 and ℓ_2 norms, respectively, for a generic J-dimensional vector, $\boldsymbol{v} = [\begin{array}{ccc} v_1 & \cdots & v_J \end{array}]'$. When $\lambda = 0$, there is no shrinkage, so that the LASSO and OLS objective functions coincide. Unlike the ℓ_2 penalty in ridge regression, the ℓ_1 penalty in Equation (3.15) permits shrinkage to zero (for sufficiently large λ), so that the LASSO performs variable selection.

The LASSO is adept at selecting relevant predictors in some settings (e.g., Zhang and Huang 2008; Bickel, Ritov, and Tsybakov 2009; Meinshausen and Yu 2009). However, it tends to arbitrarily select one predictor from a group of highly correlated predictors. The ENet (Zou and Hastie 2005) alleviates this tendency by including both ℓ_1 (LASSO) and ℓ_2 (ridge) components in the penalty term for the objection function:

$$\underset{\boldsymbol{b}\in\mathbb{R}^K}{\arg\min}\ \frac{1}{2NT}\|\boldsymbol{\delta} - \tilde{\boldsymbol{X}}\boldsymbol{b}\|_2^2 + \lambda P_\alpha(\boldsymbol{b}), \tag{3.18}$$

where

$$P_\alpha(\boldsymbol{b}) = 0.5(1-\alpha)\|\boldsymbol{b}\|_2^2 + \alpha\|\boldsymbol{b}\|_1 \tag{3.19}$$

and α is a blending parameter for the ℓ_1 and ℓ_2 components of the penalty term. When $\alpha = 1$, $P_\alpha = \|\boldsymbol{b}\|_1$ in Equation (3.19), so that the ENet reduces to the LASSO. We follow the recommendation of Hastie and Qian (2016) and set $\alpha = 0.5$.[15]

An important aspect of ENet estimation is the tuning of the regularization parameter, λ, in Equation (3.18). Tuning λ entails estimating Equation (3.18) for a grid of λ values and selecting the value that minimizes some criterion. Although K-fold cross validation is the most popular method for tuning λ, a drawback to the approach is that the number and construction of the folds are largely arbitrary. Instead, we tune the regularization parameter using information criteria. Specifically, to help to prevent overfitting in our noisy data environment, we use versions of the Bayesian information criterion (BIC, Schwarz 1978).

We begin with the conventional BIC, as suggested by Zou, Hastie, and Tibshirani (2007):

$$\mathrm{BIC} = NT \log\left(\frac{\mathrm{SSR}_\lambda}{NT}\right) + \mathrm{df}_\lambda \log(NT), \tag{3.20}$$

where SSR_λ (df_λ) is the sum of squared residuals (effective degrees of freedom) for the ENet-fitted model based on λ. Next, we consider the modified BIC proposed by Wang, Li, and Leng (2009), who modify the BIC to make it more stringent for a diverging number of predictors:

$$\mathrm{MBIC} = NT \log\left(\frac{\mathrm{SSR}_\lambda}{NT}\right) + \mathrm{df}_\lambda \log(NT) \log[\log(K)]. \tag{3.21}$$

We also use the generalized information criterion of Fan and Tang (2013), who adjust the BIC for the case where the number of predictors grows exponentially with the sample size:

$$\mathrm{GIC} = NT \log\left(\frac{\mathrm{SSR}_\lambda}{NT}\right) + \mathrm{df}_\lambda \log[\log(NT)] \log(K). \tag{3.22}$$

[15]See Hastie, Tibshirani, and Wainwright (2015) for a textbook exposition of the LASSO and ENet.

Finally, we consider a version of the extended regularization information criterion (ERIC; Hui, Warton, and Foster 2015), which refines the BIC by incorporating the value of λ:

$$\mathrm{ERIC} = NT \log\left(\frac{\mathrm{SSR}_\lambda}{NT}\right) + \mathrm{df}_\lambda \log\left(\frac{NT\hat{\sigma}^2_\lambda}{\lambda}\right), \tag{3.23}$$

where $\hat{\sigma}^2_\lambda = \mathrm{SSR}_\lambda/(NT)$.

The information criteria in Equations (3.20) to (3.23) reflect the familiar fit-parsimony tradeoff for model selection. In our context, the BIC-based criteria exert a strong shrinkage effect; in terms of the relative strength of shrinkage, the general ranking of the criteria (from least to most stringent) is BIC, GIC, MBIC, and ERIC.[16]

4 Out-of-Sample Performance

This section analyzes the accuracy of the out-of-sample forecasts. It also discusses the benefits of the pooling restriction in Equation (3.6), as well as the relevance of individual predictors as identified by the ENet.

4.1 Beating the Naïve Benchmark

Simulating the situation of a forecaster in real time, we generate exchange rate forecasts for the 1995:01 to 2019:03 out-of-sample period as follows. Reserving the first ten years of data for the initial training sample, we estimate the panel predictive regression in Equation (3.7) using data from the beginning of the available sample through 1994:12. We then use the fitted panel predictive regression and the 1994:12 predictor values for each country to compute forecasts of exchange rate changes for each available country for 1995:01. Next, we re-estimate the panel predictive regression using data through 1995:01; we then use the fitted

[16] As in Flynn, Hurvich, and Simonoff (2013) and Taddy (2017), we also consider the Hurvich and Tsai (1989) corrected version of the Akaike information criterion (AIC, Akaike 1973). However, the corrected AIC does not induce sufficient shrinkage, as it fails to outperform the naïve no-change benchmark. In addition, five-fold cross validation rarely outperforms the naïve benchmark. The complete results for five-fold cross validation and the corrected AIC are reported in Table A1 of the Internet Appendix.

panel predictive regression and the 1995:01 predictor values for each country to generate forecasts for each available country for 1995:02. We continue in this fashion through the end of the sample, which provides us with a set of exchange rate forecasts for the available countries for each of the 291 months comprising the out-of-sample period. We generate forecasts based on OLS and ENet estimation of the panel predictive regression.

For Germany, Italy, France, and the Netherlands, there are only 48 monthly forecasts (1995:01 to 1998:12) available for evaluation, due to those four countries joining the Euro area in 1999:01. After imposing a minimum requirement of twelve monthly observations before a currency is included in the panel predictive regression, there are 230 forecasts (2000:02 to 2019:03) available for the Euro area. For the remaining nine countries, forecasts are available for the entire 1995:01 to 2019:03 out-of-sample period (291 observations). In addition to reporting results for each individual country, we report results for the entire collection of forecasts taken together ($9 \times 291 + 230 + 4 \times 48 = 3,041$ observations).

To provide perspective on the degree of shrinkage induced by the different forecasting strategies, Table 2 reports the standard deviation for each of the panel predictive regression forecasts divided by the standard deviation of the realized exchange rate change for the out-of-sample period. Theoretically, monthly exchange rate changes are likely to have only a small predictable component, so that a "large" volatility ratio is a sign of overfitting. The volatility ratios for the conventional OLS forecast in the third column range from 0.31 to 0.54; the ratio for the entire set of countries is 0.41 in the last row of Table 2. Such large values suggest that the OLS forecast—which does not induce any shrinkage—misinterprets much of the noise in the data for a predictive signal. The strong shrinkage induced by the ENet forecasts is evident in the fourth through seventh columns of Table 2, as the volatility ratios are always lower than the corresponding values for the OLS forecast. With respect to the different criteria for tuning the regularization parameter, we have the following ranking in terms of the degree of shrinkage (from weakest to strongest): ENet-BIC, ENet-GIC, ENet-MBIC, and ENet-ERIC. For the ENet-ERIC forecast, the volatility ratios in the last column range from 0.03 to 0.20, which indicate that the forecast is considerably less prone to over-responding to noise.

Figure 1 provides further evidence on the degree of shrinkage induced by the various ENet forecasts. The figure shows the number of predictors selected each month by the ENet when we recursively estimate the panel predictive regression to generate the out-of-sample forecasts. According to the figure, the BIC-based criteria select a relatively limited number of predictors (20 or less) each month, so that the fitted panel predictive regressions are fairly sparse. The ENet-ERIC nearly always selects the most parsimonious model.

The relatively strong shrinkage effect exerted by the ENet-ERIC forecast can be directly seen in Figure 2, which depicts the OLS and ENet-ERIC forecasts, along with the actual exchange rate change.[17] In line with the volatility ratios in the third and seventh columns of Table 2, Figure 2 shows that the ENet-ERIC approach substantially stabilizes the exchange rate forecasts vis-à-vis the conventional OLS approach. The more extreme forecasts produced by OLS estimation of the panel predictive regression are clearly evident in Figure 2 and point to substantive overfitting; ENet-ERIC estimation induces considerable shrinkage in the forecast, thereby helping to guard against overfitting.

Next, we assess the accuracy of the exchange rate forecasts in terms of MSFE. We can conveniently compare the relative accuracy of the forecast based on the panel predictive regression in Equation (3.14) to the naïve no-change benchmark forecast using the Campbell and Thompson (2008) out-of-sample R^2 statistic:

$$R^2_{\text{OS}} = 1 - \frac{\sum_{t=T_1+1}^{T_2} \left(\hat{e}^{\text{PPR}}_{i,t|t-1} \right)^2}{\sum_{t=T_1+1}^{T_2} \left(\hat{e}^{\text{bench}}_{i,t|t-1} \right)^2}, \tag{4.1}$$

where

$$\hat{e}^{\text{PPR}}_{i,t|t-1} = \delta_{i,t} - \hat{\delta}^{\text{PPR}}_{i,t|t-1}, \tag{4.2}$$

[17] Corresponding figures for the ENet-BIC, ENet-MBIC, and ENet-GIC forecasts are provided in Figures A1 through A3 of the Internet Appendix.

$$\hat{e}^{\text{bench}}_{i,t|t-1} = \delta_{i,t} - \hat{\delta}^{\text{bench}}_{i,t|t-1}, \tag{4.3}$$

$\hat{\delta}^{\text{bench}}_{i,t|t-1} = 0$ is the no-change benchmark forecast, T_1 is the last observation for the initial in-sample period, and T_2 is the last available observation.[18] The R^2_{OS} statistic measures the proportional reduction in MSFE for a competing forecast vis-à-vis the benchmark. Because the predictable component in monthly exchange rate changes is inherently small, the R^2_{OS} statistic will necessarily be small. Nevertheless, even a seemingly small degree of exchange rate predictability can be economically meaningful, as we show in Section 5 below and as is the case for stock return predictability (Campbell and Thompson 2008). To get a sense of whether the competing forecast provides a statistically significant improvement in MSFE relative to the benchmark, we compute the Clark and West (2007) adjusted version of the Diebold and Mariano (1995) and West (1996) statistic.[19]

Table 3 reports R^2_{OS} statistics for the OLS and ENet forecasts. The R^2_{OS} statistics are all negative in the third column of Table 3, so that the OLS forecast always fails to outperform the no-change benchmark in terms of MSFE. The negative R^2_{OS} statistics are often sizable in magnitude for the individual countries, and the statistic is -8.03% for all of the countries taken together in the last row. The overfitting in the OLS forecast detected in Table 2 and Figures 1 and 2 thus translates into poor out-of-sample performance when it comes to forecast accuracy in Table 3.

As shown in the fourth through seventh columns of Table 3, the ENet forecasts evince markedly better out-of-ample performance. For the ENet-BIC forecast in the fourth column, the R^2_{OS} statistics are positive for eight of the G10 countries (which have at least 230 available out-of-sample observations); the R^2_{OS} statistics are significant at conventional levels for all eight of those counties. The positive R^2_{OS} statistics range from 1.30% to 4.79%, which are

[18] For our application, T_1 and T_2 correspond to 1994:12 and 2019:03, respectively.

[19] As shown by Clark and McCracken (2001) and McCracken (2007), the Diebold-Mariano-West statistic has a non-standard asymptotic distribution when comparing forecasts from nested models (as is the case for our application). In particular, the Diebold-Mariano-West statistic can be severely undersized when comparing nested forecasts, meaning that it can have little power to detect a significant improvement in forecast accuracy.

quite large in the context of the literature surveyed by Rossi (2013). For the complete set of countries in the last row, the R^2_{OS} statistic is 1.72% (significant at the 1% level).

The ENet-MBIC and ENet-GIC forecasts (see the fifth and sixth columns, respectively, of Table 3) both produce positive R^2_{OS} statistics for ten of the 14 countries, as well as eight of the G10 countries; for the latter group, the R^2_{OS} statistics are significant for seven (all eight) of the countries with positive R^2_{OS} statistics for the ENet-MBIC (ENet-GIC) forecast. Many of the R^2_{OS} statistics are sizable, reaching as high as 4.09% (4.85%) for the ENet-MBIC (ENet-GIC) forecast in the case of the United Kingdom. For all of the countries taken together, the ENet-MBIC and ENet-GIC forecasts generate R^2_{OS} statistics of 1.38% and 1.60%, respectively, both of which are significant at the 1% level.

The ENet-ERIC forecast in the final column of Table 3 displays the best overall performance: 13 of the 14 R^2_{OS} statistics are positive for the individual countries, including all ten of the R^2_{OS} statistics for the G10 countries; eight of those statistics are significant at conventional levels. Many of the R^2_{OS} statistics are again sizable—seven are greater than or equal to 1.95%, including 5.32% for the United Kingdom. For the entire group of countries, the R^2_{OS} statistic is 2.04% (significant at the 1% level) for the ENet-ERIC forecast, which is the highest value in the last row of Table 3.

Overall, Table 3 demonstrates that the ENet is an effective machine learning device for forecasting exchange rates using large information sets, provided that we use stringent BIC-based information criteria to tune the regularization parameter. The ENet-ERIC forecast—which employs the most stringent criterion—consistently outperforms the no-change benchmark and provides among the most accurate short-horizon exchange rate forecasts available to date.

To examine the performance of the ENet-ERIC forecast over time, Figure 3 employs the graphical device of Goyal and Welch (2003, 2008). The figure portrays the cumulative difference in squared forecast errors for the no-change benchmark vis-à-vis the ENet-ERIC forecast.[20] Each curve conveniently allows for a comparison of forecast accuracy (in terms of MSFE) for any subsample: we compare the height of the curve at the beginning and end of the segment corresponding to the subsample; if the curve is higher (lower) at the end of the segment, then the ENet-ERIC (no-change) forecast has a lower MSFE for the subsample. A forecast that always outperforms the benchmark will have a curve with a uniformly positive slope. Of course, given that exchange rate changes have a large unpredictable component, this ideal is unattainable in practice. Realistically, we seek a forecast with a curve that is predominantly positively sloped and that does not have extended segments with negative slopes.

According to Figure 3, the curves are positively sloped for much of the time for the United Kingdom, Switzerland, Australia, New Zealand, Sweden, Norway, Denmark, the Euro area, Germany, and France, so that the ENet-ERIC forecast generates out-of-sample gains on a consistent basis over time. The gains are less consistent over time for Japan, Canada, and the Netherlands. A striking feature of Figure 3 is the improvement in accuracy provided by the ENet-ERIC forecast vis-à-vis the no-change benchmark during the worst phase of the global financial crisis in late 2008. As we show in Section 5, this has important implications for carry trade investment strategies.

4.2 Benefits From Pooling with Many Predictors

The panel predictive regression in Equation (3.6) imposes the pooling restriction that the slope coefficients are homogeneous across countries. From an out-of-sample forecasting perspective, imposing such a restriction is potentially beneficial in light of the bias-variance tradeoff: although the pooling restriction likely introduces bias in the coefficient estimates, this can be offset by the gain in precision associated with having fewer parameters to estimate. The precision gain is germane in our high-dimensional and noisy data environment. In essence, pooling helps to further alleviate overfitting.

[20] Corresponding figures for the ENet-BIC, ENet-MBIC, and ENet-GIC forecasts are provided in Figures A4 through A6 of the Internet Appendix.

We find that pooling substantially improves out-of-sample exchange rate forecasts based on our set of 70 predictors. As shown in Table A2 of the Internet Appendix, forecasts based on predictive regressions estimated at the country level (i.e., without imposing the slope homogeneity restriction) rarely outperform the no-change benchmark, whether the regressions are estimated via OLS or the ENet. In no case is the R^2_{OS} statistic higher than the corresponding value for the ENet-ERIC forecast in the last column of Table 3.

To determine if it is beneficial to use the information in a large number of predictors, we generate out-of-sample forecasts based on univariate predictive regressions, where each of the ten individual country characteristics appear in turn as predictors.[21] In addition, we generate forecasts based on a predictive regression with INF and BILL appearing jointly as predictors, in line with Taylor rule fundamentals. The results, reported in Tables A3 and A4 of the Internet Appendix, reveal that it is advantageous to utilize a large information set when forecasting exchange rates (provided that we guard against overfitting): forecasts based on individual characteristics and Taylor rule fundamentals typically fail to outperform the no-change benchmark, and the R^2_{OS} statistics are rarely larger than the corresponding values for the ENet-ERIC forecast in Table 3.

In sum, by imposing the slope homogeneity restriction and estimating a panel predictive regression with 70 predictors via the ENet (and using a stringent criterion to tune the regularization parameter), we can utilize the information in a large number of potentially relevant predictors to consistently outperform the naïve no-change forecast, which has proven stubbornly difficult to beat, especially at the monthly horizon.

4.3 Which Predictors Matter?

Because it performs variable selection, the ENet is a machine learning device that facilitates model interpretation in high-dimensional settings. We investigate the relevance of the individual predictors through the lens of the ENet-ERIC recursive slope coefficient estimates.

[21] We also use the difference between the log CPIs for country i and the United States, in accord with purchasing power parity.

To shed light on which predictors are selected in the fitted panel predictive regressions on average over time, Figure 4 provides a heatmap for the selection frequencies for the individual predictors selected by the ENet-ERIC.[22]

The first column of Figure 4 indicates that, among the country characteristics on their own (i.e., not interacted with the global variables), DP is selected much of the time, followed by PE and IV. According to the remaining columns, a number of the predictors based on the interactions of the country characteristics with the global variables are relevant for predicting exchange rates. In particular, predictors based on the interactions of INF, UN, BILL, and NOTE with GVOL appear germane. The interaction of SRET with MPU also appears important. In addition, the interactions of DP with EPU; PE with EPU, GR, and GILL; and UN, DP, and IV with GCOR are selected fairly frequently. From a practical standpoint, we cannot know ex ante which of 70 predictors are the most relevant. The ENet-ERIC provides a data-driven method for selecting the relevant predictors from a large set of potential predictors.

Figure 5 depicts the recursive OLS and ENet-ERIC coefficient estimates for the panel predictive regression.[23] The figure illustrates the strong shrinkage effect induced by ENet-ERIC estimation. The OLS estimates are highly volatile; indeed, a given OLS coefficient estimate often changes sign over time. Such behavior reflects the substantial imprecision associated with conventional OLS estimation in our high-dimension setting, a manifestation of overfitting. In contrast, the ENet-ERIC coefficient estimates are much more stable over time, and the nonzero estimates never change sign for a given predictor. As expected, the ENet-ERIC coefficient estimates are often considerably smaller in magnitude than their OLS counterparts.

INF, UN, and BILL are popular predictors in the exchange rate literature. The first column in Figure 4 shows that these predictors are rarely selected on their own; however,

[22]Heatmaps for the selection frequencies for the ENet-BIC, ENet-MBIC, and ENet-GIC are provided in Figures A7 through A9 of the Internet Appendix.

[23]Corresponding figures for the ENet-BIC, ENet-MBIC, and ENet-GIC are available as Figures A10 through A12 of the Internet Appendix.

as shown in the fifth column of Figure 4, they are frequently selected when interacted with GVOL. The ENet-ERIC recursive coefficient estimates for the interactions of INF, UN, and BILL with GVOL (INF.GVOL, UN.GVOL, and BILL.GVOL, respectively) in Figure 5 reveal an interesting pattern. According to Taylor rule logic, an increase (decrease) in INF (UN) leads to an increase in expected US dollar appreciation. The nonzero ENet-ERIC recursive coefficient estimates for INF.GVOL (UN.GVOL) are always positive (negative), so that the expected US dollar appreciation in response to an increase in INF (UN) becomes larger (smaller) as GVOL increases; in other words, as global foreign exchange volatility increases, Taylor rule predictions become more relevant. Similarly, the nonzero ENet-ERIC recursive coefficient estimates for BILL.GVOL are uniformly positive. UIP predicts that an increase in BILL corresponds to an increase in expected dollar appreciation, meaning that the relationship predicted by UIP holds to a greater degree as global foreign exchange volatility increases (as it does, e.g., during the global financial crisis).

An emerging literature (e.g., Engel and Wu 2019; Jiang, Krishnamurthy, and Lustig 2019; Kremens and Martin 2019; Adrian and Xie 2020; Lilley et al. forthcoming) finds evidence of exchange rate predictability around the global financial crisis based on the US dollar's perception as a safe-haven currency. Due to data availability, these studies analyze medium- to long-horizon predictability (i.e., at a quarterly horizon or longer) and/or employ relatively short samples. Although we focus on short-horizon predictability and consider a longer out-of-sample period, our ENet-ERIC forecast appears to capture exchange rate predictability around the crisis relating to the US dollar's safe-haven status. Specifically, the forecasts in Figure 2 portend a relatively strong depreciation for many countries' currencies in late 2008 during the worst phase of the crisis; as shown in Figure 3, the ENet-ERIC forecast substantially outperforms the no-change benchmark during that time. To the extent that the out-of-sample gains around the crisis are due to a dash to the US dollar resulting from a decline in global risk tolerance, the predictors selected by the ENet-ERIC appear to capture the effect.[24]

We also examine links between the set of predictors selected by the ENet-ERIC and the Lilley et al. (forthcoming) capital flow measure, which is available at the quarterly frequency and is based on the change in US holdings of foreign bonds. Lilley et al. (forthcoming) find that their measure is significantly related to various proxies of global risk appetite, as well as the change in a broad US dollar index from 2007 to 2017, which they interpret as evidence of the US dollar's role as a safe-haven currency during the crisis. We investigate links between the predictors selected by the ENet-ERIC and the change in US foreign bond holdings. We begin with the group of ten predictors selected by the ENet-ERIC from the last set of recursive coefficient estimates in Figure 5 (which uses the largest sample size).[25] We then average the predictors over the three months comprising a quarter, as well as across countries. Finally, we regress the change in US foreign bond holdings on the set of ten time- and country-aggregated predictors for 2007:1 to 2019:2.[26]

Figure 6 shows the (standardized) change in US foreign bond holdings, together with the fitted values for the regression. As a group, the predictors selected by the ENet-ERIC are significantly related to the change in US foreign bond holdings at the 1% level, and the predictors collectively explain 44.33% of the variation in the capital flow measure. The fitted values track the actual values quite closely during the crisis, providing further evidence that the relevant predictors identified by the ENet-ERIC contain information pertaining to the US dollar's safe-haven status.

[24]Section 5.3 provides additional evidence in this regard.

[25]The ten predictors are DP, PE, INF.GVOL, UN.GVOL, UN.GCOR, BILL.GVOL, DP.GCOR, PE.GILL, SRET.MPU, and IV.GCOR.

[26]Data for the change in US foreign bond holdings from Lilley et al. (forthcoming) are available from the Global Capital Allocation Project website.

5 Carry Trade Portfolios

In this section, we analyze the economic value of the ENet forecasts in the context of carry trade portfolios.

5.1 Portfolio Construction

We consider a US investor with mean-variance preferences who allocates funds across all available foreign currencies. At the end of T, the investor's objective function is given by

$$\underset{\boldsymbol{w}_{T+1}}{\arg\min}\ \boldsymbol{w}'_{T+1}\hat{\boldsymbol{\mu}}_{T+1|T} - 0.5\gamma\boldsymbol{w}'_{T+1}\hat{\boldsymbol{\Sigma}}_{T+1|T}\boldsymbol{w}_{T+1}, \tag{5.1}$$

where

$$\hat{\boldsymbol{\mu}}_{T+1|T} = [\ \widehat{RX}_{1,T+1|T}\ \ \cdots\ \ \widehat{RX}_{N,T+1|T}\ ]', \tag{5.2}$$

$$\widehat{RX}_{i,T+1|T} = (r_{i,T} - r_{\text{US},T}) - \hat{\delta}_{i,T+1|T} \tag{5.3}$$

is the investor's excess return forecast for country i's currency, $\hat{\delta}_{i,T+1|T}$ generically denotes an exchange rate change forecast, $\hat{\boldsymbol{\Sigma}}_{T+1|T}$ is the investor's estimate of the variance-covariance matrix for the currency excess returns, $\boldsymbol{w}_{T+1} = [\ w_{1,T+1}\ \ \cdots\ \ w_{N,T+1}\ ]'$ is the N-vector of portfolio weights that will be in effect for $T+1$, and γ is the coefficient of relative risk aversion.

We consider two cases, which differ with respect to the exchange rate forecast used to compute $\widehat{RX}_{i,T+1|T}$ in Equation (5.3). In the first case, the investor uses the no-change benchmark forecast:

$$\widehat{RX}^{\text{bench}}_{i,T+1|T} = (r_{i,T} - r_{\text{US},T}) - \underbrace{\hat{\delta}^{\text{bench}}_{i,T+1|T}}_{0} = r_{i,T} - r_{\text{US},T}. \tag{5.4}$$

For this case, the investor assumes that the exchange rate change is unpredictable, so that the currency excess return forecast coincides with the bill yield differential ($\text{BILL}_{i,t}$, which is known at T). As is common among practitioners, we assume that the investor uses an exponentially weighted moving average (EWMA) estimator for $\hat{\mathbf{\Sigma}}_{T+1|T}$.[27] Furthermore, we assume that $\gamma = 5$; the results are qualitatively similar for reasonable alternative γ values. Finally, to keep each portfolio weight in a plausible range, we impose the restriction that $-0.5 \leq w_{i,T+1} \leq 0.5$ for $i = 1, \ldots, N$. The carry trade portfolio based on Equation (5.4), Basic-Opt, is the optimal portfolio for the mean-variance investor when they simply use the bill yield differential to forecast each currency excess return.

For the second case, the investor uses the exchange rate forecast in Equation (3.14) based on ENet-ERIC estimation of the panel predictive regression to compute the currency excess return forecast:

$$\widehat{RX}^{\text{PPR}}_{i,T+1|T} = (r_{i,T} - r_{\text{US},T}) - \hat{\delta}^{\text{PPR}}_{i,T+1|T}. \tag{5.5}$$

For this case, the investor assumes that the exchange rate change is predictable and relies on the ENet-ERIC forecast. Otherwise, we again assume that the investor uses an EWMA estimator for $\hat{\mathbf{\Sigma}}_{T+1|T}$, sets γ equal to five, and imposes the portfolio weight restriction. The carry trade portfolio based on Equation (5.5), Smart-Opt, seeks to improve upon the Basic-Opt portfolio by using the ENet-ERIC forecast to refine the currency excess return forecast.[28]

We construct out-of-sample portfolio weights as follows. We first use data through 1994:12 to compute the EWMA estimate of the variance-covariance matrix and ENet-ERIC exchange rate forecasts for 1995:01. We then solve Equation (5.1) using the no-change (ENet-ERIC) exchange rate forecasts to compute the Basic-Opt (Smart-Opt) portfolio weights for 1995:01. Next, we use data through 1995:01 to generate the EWMA variance-covariance matrix esti-

[27]We use a value of 0.94 for the decay parameter in the EWMA variance-covariance matrix estimator, a value often used in practice.

[28]The results are similar for Smart-Opt portfolios based on the ENet-BIC, ENet-MBIC, and ENet-GIC forecasts. The complete results are reported in Tables A5 through A7 of the Internet Appendix.

mate and ENet-ERIC forecasts for 1995:02; we then compute the Basic-Opt and Smart-Opt portfolio weights for 1995:02. We proceed in this fashion through the end of the out-of-sample period, so that we simulate the situation of an investor in real time.[29] By comparing the performance of the Basic-Opt portfolio (which assumes that the exchange rate change is not predictable) to that of the Smart-Opt portfolio (which relies on the ENet-ERIC forecast of the exchange rate change), we can gauge the economic value of exchange rate predictability for an investor.

5.2 Portfolio Performance

Table 4 reports annualized means, volatilities, and Sharpe ratios for the Basic-Opt and Smart-Opt portfolio excess returns. In additional to the full 1995:01 to 2019:03 out-of-sample period, the table reports results for the 1995:01 to 2008:08 and 2008:09 to 2019:03 subsamples, which we refer to as the pre- and post-crisis subsamples, respectively. The start of the second subsample coincides with the bankruptcy of Lehman Brothers on September 15, 2008 at the height of the global financial crisis.

We report results for different assumptions regarding transaction costs. First, we ignore transaction costs by using mid-quotes for the bid-ask spreads from Datastream for the relevant forward and spot exchange rates to compute the currency excess returns. We also compute portfolio excess returns using the bid and ask quotes from Datastream for the relevant forward and spot rates (e.g., Lustig, Roussanov, and Verdelhan 2011), thereby assuming that the investor has to pay the bid and ask prices reported in Datastream. A number of studies document that the bid-ask spreads offered by Datastream are unrealistically high (e.g., Lyons 2001; Neely, Weller, and Ulrich 2009; Menkhoff et al. 2012b; Neely and Weller 2013). Specifically, investors trade the best quoted price at each point in time, making the full spread in Datastream considerably higher than the effective spread for foreign exchange

[29] Equations (1.1) and (5.3) are approximately equal to the excess return corresponding to US dollar accumulation, which is given by $RX_{i,t+1} = \left(S^d_{i,t+1} - F^d_{i,t}\right)/S^d_{i,t}$, where $S^d_{i,t} = 1/S_{i,t}$ and $F^d_{i,t} = 1/F_{i,t}$. We use the relevant forward and spot exchange rates to compute the exact excess return for each foreign currency when calculating the Basic-Opt and Smart-Opt portfolio excess returns for a US investor.

market participants.[30] To more accurately reflect the relevant transaction costs faced by traders, we follow Goyal and Saretto (2009) and Menkhoff et al. (2012b) and also report results for currency excess returns computed using 25%, 50%, and 75% of the quoted bid-ask spreads from Datastream.

Panel A of Table 4 reports performance measures for the full out-of-sample period. The second through fourth columns provide results for the Basic-Opt portfolio. Ignoring transaction costs, the annualized mean and volatility for the portfolio excess return are 7.47% and 10.22%, respectively, which yield an impressive annualized Sharpe ratio of 0.73 (significant at the 1% level). For reasonable transaction costs of 25% and 50% of the quoted bid-ask spreads, the Sharpe ratios remain sizable at 0.56 and 0.49, respectively (significant at the 1% and 5% levels, respectively); for the full bid-ask spread, the Sharpe ratio declines to 0.34 (significant at the 10% level).

The results for the Basic-Opt portfolio in Panel A of Table 4 mask stark differences in the portfolio's performance over time. For the first subsample in Panel B, the average returns and Sharpe ratios are considerably higher than those for the full sample, with the average returns (Sharpe ratios) ranging from 6.83% to 11.88% (0.64 to 1.13, significant at the 5% or 1% levels) for the different assumptions regarding transaction costs. As shown in Panel C, beginning in September of 2008, the average returns and Sharpe ratios decline dramatically, ranging from -1.04% to 1.76% and -0.11 to 0.18, respectively; none of the Sharpe ratios are significant at conventional levels for the 2008:09 to 2019:03 subsample.

We also construct a conventional carry trade portfolio that sorts currencies into quintiles according to interest rate differentials and then takes equally weighted long (short) positions in the currencies in the fifth (first) quintile. This is tantamount to the carry trade risk factor in Lustig, Roussanov, and Verdelhan (2011). The conventional carry portfolio generally does not perform as well as the Basic-Opt portfolio; for example, ignoring transaction costs, the

[30] The foreign exchange market is one of the most liquid markets, with low transaction costs and no natural short-selling constraints. According to the 2016 Bank for International Settlements Triennial Survey, average daily turnover in the foreign exchange market is five trillion US dollars.

conventional portfolio generates annualized Sharpe ratios of 0.49, 0.75, and 0.23 for the full out-of-sample period and pre- and post-crisis subsamples, respectively.[31]

The fifth through seventh columns of Table 4 report performance metrics for the Smart-Opt portfolio, which utilizes the ENet-ERIC exchange rate forecasts when computing the currency excess return forecasts. For the full out-of-sample period, in the absence of transaction costs, the annualized average return and volatility are 12.94% and 12.80%, respectively, which translate into a quite sizable annualized Sharpe ratio of 1.01 (significant at the 1% level). With transaction costs of 25% and 50% of the quoted bid-ask spread, the Sharpe ratios remain large (0.82 and 0.76, respectively, both of which are significant at the 1% level). The results for the Smart-Opt portfolio for the pre-crisis subsample in Panel B are similar to those for the Basic-Opt portfolio, while they differ markedly for the post-crisis subsample in Panel C. Unlike the Basic-Opt portfolio, the Smart-Opt portfolio continues to deliver sizable average returns and Sharpe ratios in Panel C. Ignoring transaction costs, the Sharpe ratio for the Smart-Opt portfolio is 0.87 (significant at the 1% level) for the post-crisis subsample; the Sharpe ratios range from 0.63 to 0.75 (all of which are significant at the 5% level) for the various assumptions regarding transaction costs.

Overall, Table 4 shows that the Basic-Opt portfolio performs impressively through the summer of 2008; however, it subsequently suffers a pronounced deterioration in performance. By utilizing the ENet-ERIC forecasts, the Smart-Opt portfolio delivers impressive performance consistently over time.

To provide additional perspective on relative performance, Figure 7 depicts log cumulative excess returns for the Basic-Opt, Smart-Opt, and conventional carry trade portfolios. The Basic-Opt and Smart-Opt portfolios perform similarly through the summer of 2008, although the Smart-Opt portfolio fares better in the wake of the Asian financial and Long-Term Capital Management crises in 1998. While both portfolios experience losses in September of 2008 during the Lehman bankruptcy, their subsequent performances differ markedly, in line with

[31] The complete results are available in Table A8 of the Internet Appendix.

Panel C of Table 4. The Basic-Opt portfolio suffers more sizable losses later in 2008 in Figure 7, and its cumulative return essentially "flatlines" thereafter. The conventional carry portfolio, which is not based on an optimization framework, suffers an even larger drop in late 2008 compared to the Basic-Opt portfolio and also flatlines subsequently. In contrast, the Smart-Opt portfolio makes a strong recovery in late 2008 and continues to produce gains thereafter. The relatively strong performance of the Smart-Opt portfolio in late 2008 and subsequently translates into a substantive gain in terminal wealth: for the Smart-Opt portfolio, a US investor who begins with $1 at the end of 1994:12 ends up with $18.76 at the end of 2019:03, compared to only $5.35 for the Basic-Opt portfolio (and $2.87 for the conventional carry portfolio).

The relatively strong performance of the Smart-Opt portfolio in late 2008 also aligns with the large out-of-sample gains accruing to the ENet-ERIC forecast vis-à-vis the no-change benchmark during that time in Figure 3. The Basic-Opt portfolio simply uses the bill yield differential to forecast the currency excess return—in line with the no-change benchmark forecast—while the Smart-Opt portfolio incorporates the information in the predictors via the ENet-ERIC forecast. The statistical gains generated by the ENet-ERIC forecast relative to the naïve benchmark in late 2008 and beyond in Figure 3 translate into economic gains in the form of improved portfolio performance in Figure 7.

Further evidence on the links between exchange rate predictability and the carry portfolios is furnished by Figure 8, which portrays the currency weights for the Basic-Opt and Smart-Opt portfolios. As the figure illustrates, the ENet-ERIC forecasts often lead to substantially different allocations. The differences in allocations produced by the ENet-ERIC forecast vis-à-vis the no-change benchmark in Equation (5.1) deliver improved carry trade portfolio performance in Table 4 and Figure 7.

5.3 A Closer Look at Late 2008

Figure 7 shows that late 2008 is a perilous time for the Basic-Opt portfolio (as well as the conventional carry portfolio). Figures 9 and 10 provide additional insight into the sources of the poor performance of the Basic-Opt portfolio in late 2008, as well as how the Smart-Opt portfolio improves performance. Figure 9 shows the currency excess return forecasts that serve as inputs in the portfolio optimization problem in Equation (5.1), along with the actual excess returns, for September through December of 2008. Because the Basic-Opt portfolio uses the no-change exchange rate forecast, the benchmark currency excess return forecast is simply the bill yield differential; the Smart-Opt portfolio augments the bill yield differential with the ENet-ERIC forecast of the exchange rate change, as in Equation (5.5). Figure 10 displays the currency weights for the portfolios.

In September of 2008, the benchmark and ENet-ERIC currency excess return forecasts in Panel A of Figure 9 lead to allocations of the same sign (and typically similar magnitude) in Panel A of Figure 10 for Switzerland, Canada, Australia, New Zealand, Sweden, Norway, and Denmark. The ENet-ERIC currency excess return forecasts generate notable differences in allocations for the other countries: for the United Kingdom and Euro area, the Basic-Opt (Smart-Opt) portfolio takes long (short) positions; for Japan, the Basic-Opt (Smart-Opt) portfolio takes a short (long) position. As shown in Panel A of Figure 9, with the exception of Japan, all of the actual currency excess returns are negative in September 2008, consistent with the US dollar (as well as the Japanese yen) being viewed as a safe-haven currency; the negative returns are large in magnitude for Australia, New Zealand, Sweden, Norway, Denmark, and the Euro area. On the basis of the allocations and actual excess returns, the Basic-Opt portfolio suffers a loss of -8.02% in September of 2008, while the loss is less than half as large (-3.84%) for the Smart-Opt portfolio. The differences in allocations signaled by the ENet-ERIC forecast—especially for the United Kingdom, Japan, and the Euro area—limit portfolio losses in the month of the Lehman bankruptcy.

The currency excess return forecasts differ sharply during October of 2008 in Panel B of Figure 9: with the exception of Japan, all of the benchmark currency excess return forecasts are positive, while the ENet-ERIC forecasts are negative for all ten of the countries. The differences in currency excess return forecasts correspond to a number of markedly different allocations in Panel B of Figure 10; most notably, the Basic-Opt (Smart-Opt) portfolio exhibits sizable long (short) positions for the United Kingdom, New Zealand, Norway, Denmark, and the Euro area. With the exception of Japan, all of the actual currency excess returns are again negative in Panel B of Figure 9, and the negative returns are larger in magnitude than those in Panel A.[32] The different allocations prompted by the ENet-ERIC forecast vis-à-vis the no-change benchmark enable the Smart-Opt portfolio to substantively outperform the Basic-Opt portfolio during October of 2008: the latter suffers a loss of −13.53%, while the former enjoys a massive gain of 27.01%. The situation in terms of the currency excess return forecasts and allocations is similar in November of 2008 in Panel C of Figures 9 and 10. Because fewer of the currency excess returns are negative and most are smaller in magnitude, the Basic-Opt experiences a gain of 1.53% during the month; however, the information in the ENet-ERIC forecasts leads to a much larger gain of 14.22% for the Smart-Opt portfolio.

The environment appears to normalize to an extent in December of 2008, as the discrepancies between the benchmark and ENet-ERIC forecasts in Panel D of Figure 9, as well as those between the Basic-Opt and Smart-Opt portfolio weights in Panel D of Figure 10, are more muted, although important differences remain (e.g., the portfolio weights for Switzerland). In addition, with the exception of the United Kingdom, all of the currency excess returns are positive in December of 2008. The Basic-Opt portfolio realizes a gain of 5.56% for the month, while the gain for the Smart-Opt portfolio is more than twice as large (12.01%).[33]

[32] Note the difference in the scales for the vertical axes across Panels A and B.

[33] For September through December of 2008, the excess returns for the conventional carry portfolio are −5.96%, −13.57%, −2.99%, and −5.54%, in line with the sharp drop in the portfolio's cumulative excess return in Figure 7.

Figures 9 and 10 help to explain how exchange rate predictability—as captured by our ENet-ERIC forecast—is especially valuable to an investor during the worst part of the global financial crisis in late 2008. By anticipating a depreciation in many foreign currencies, the ENet-ERIC forecast leads to sizable negative positions in many currencies, enabling the Smart-Opt portfolio to avoid the large losses suffered by a basic carry strategy and even realize large gains. In essence, the Smart-Opt portfolio shorts the traditional carry strategy to a significant extent during the tumult of the global financial crisis.

5.4 Alphas

Next, we examine whether the Basic-Opt and Smart-Opt portfolios generate alpha in the context of the Lustig, Roussanov, and Verdelhan (2011) currency factor model. The model includes dollar and carry trade risk factors, denoted by $\mathrm{MKT}_{\mathrm{FX}}$ and $\mathrm{HML}_{\mathrm{FX}}$, respectively. The dollar factor is an equally weighted average of the available currency excess returns for the month, while the carry trade risk factor is the conventional carry portfolio defined previously. Table 5 reports factor model estimation results for the Basic-Opt and Smart-Opt portfolios. We again report results for the full out-of-sample period, as well as the pre- and post-crisis subsamples.

For the full 1995:01 to 2019:03 out-of-sample period, the Smart-Opt portfolio generates a large annualized alpha of 10.83% in the sixth column, which is almost three times greater than that for the Basic-Opt portfolio (3.82%) in the second column. The former (latter) is significant at the 1% (5%) level. The more limited alpha for the Basic-Opt portfolio is due to its essentially zero exposure to the dollar factor and substantive exposure to the carry factor (0.75, significant at the 1% level). The Smart-Opt portfolio evinces exposures of -0.52 and 0.48 to the dollar and carry factors, respectively, both of which are significant at the 1% level. The two factors together explain substantially more of the return variation for the Basic-Opt vis-à-vis the Smart-Opt portfolio (adjusted R^2 statistics of 52.19% and 16.90%, respectively).

Table 5 reveals important differences in performance for the Basic-Opt portfolio across the two subsamples. For the pre-crisis subsample, the Basic-Opt portfolio exhibits near unitary exposure to the carry trade factor (0.96, significant at the 1% level), and it generates a sizable annualized alpha of 5.17% (significant at the 1% level). Reminiscent of Table 4 and Figure 7, the Basic-Opt portfolio's performance deteriorates sharply during the post-crisis subsample. It continues to display substantial exposure to the carry factor (0.59, significant at the 1% level), while its annualized alpha declines to only 0.25% (insignificant at conventional levels). In contrast, the Smart-Opt portfolio delivers impressive annualized alphas of 8.54% and 11.41% for the first and second subsamples, respectively (significant at the 1% and 5% levels, respectively). While the Smart-Opt portfolio exhibits a sizable exposure of 0.88 to the carry factor (significant at the 1% level) for the first subsample, its exposure is only 0.05 (insignificant at conventional levels) for the second subsample. The information contained in the ENet-ERIC forecast thus leads the investor to effectively disconnect from a conventional carry strategy in the second subsample.[34]

6 Deep Learning

To this point, we permit a limited degree of nonlinearity in the predictive regression via the interaction terms involving the country characteristics multiplied by the global variables; however, the panel predictive regression, Equation (3.6), remains linear. In this section, we allow for more complex predictive relationships through artificial neural networks (or nets), a popular machine learning tool for prediction.

6.1 Architecture and Training

We consider *feedforward* neural nets, the most well-known type of neural net. Gu, Kelly, and Xiu (2020) recently find that feedforward neural nets are useful for improving time-series

[34]Tables A9 and A10 of the Internet Appendix report results for Basic-Opt and Smart-Opt portfolios for non-US domestic investors. The results are qualitatively similar to those for a US domestic investor.

forecasts of individual stock returns using firm characteristics (as well as firm characteristics interacted with macroeconomic variables). We explore whether neural nets are useful for forecasting exchange rate changes based on our set of 70 predictors.

A neural net architecture is comprised of multiple layers. The first is the *input* layer, which is simply the set of predictors, which we denote by $x_1, \dots, x_{K_0}$. One or more *hidden* layers are next. Each hidden layer l contains K_l neurons, each of which takes predictive signals from the neurons in the previous hidden layer to produce another signal:

$$h_m^{(l)} = g\left(\omega_{m,0}^{(l)} + \sum_{j=1}^{K_{l-1}} \omega_{m,j}^{(l)} h_j^{(l-1)}\right) \text{ for } m = 1, \dots, K_l;\ l = 1, \dots, L, \tag{6.1}$$

where $h_m^{(l)}$ is the signal corresponding to the mth neuron in the lth hidden layer;[35] $\omega_{m,0}^{(l)}, \omega_{m,1}^{(l)}, \dots, \omega_{m,K_{l-1}}^{(l)}$ are weights ($\omega_{m,0}^{(l)}$ is known as the bias term); and $g(\cdot)$ is an activation function. The *output* layer is an affine function that translates the signals from the last hidden layer into a prediction:

$$\hat{\delta} = \omega_0^{(L+1)} + \sum_{j=1}^{K_L} \omega_j^{(L+1)} h_j^{(L)}, \tag{6.2}$$

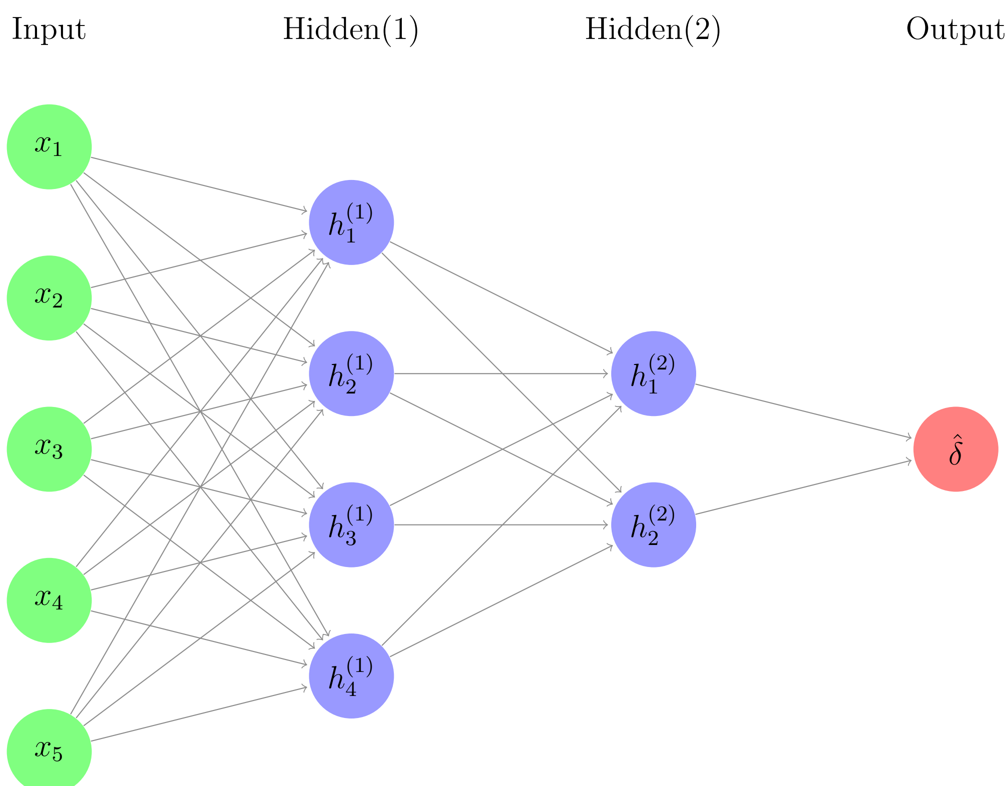
where $\hat{\delta}$ denotes the forecast of the target variable. For the activation function, we use the popular rectified linear unit (ReLU) function:[36]

$$g(x) = \begin{cases} 0 & \text{if } x < 0, \\ x & \text{otherwise.} \end{cases} \tag{6.3}$$

Intuitively, the activation function activates a neuronal connection in response to a sufficiently strong signal, thereby relaying the signal forward through the network.

[35] For the first hidden layer, $h_j^{(0)} = x_j$ for $j = 1, \dots, K_0$.
[36] Gu, Kelly, and Xiu (2020) also use the ReLU activation function

The following diagram provides a schematic for a relatively simple feedforward neural net architecture with five predictors and two hidden layers, where the hidden layers contain four and two neurons, respectively. The diagram shows that the five predictors in the input layer feed through to provide signals to each of the four neurons in the first hidden layer; the neurons in the first hidden layer subsequently feed through to provide signals to each of the two neurons in the second hidden layer. The neurons in the second hidden layer provide a final set of signals for the output layer. There are $(5+1)4 = 24$ weights for the first hidden layer, $(4+1)2 = 10$ weights for the second layer, and $2+1 = 3$ weights for the output layer, for a total of 37 weights.



There is substantial judgment involved in specifying a neural net architecture. Theoretically, a neural net with a single hidden layer and sufficiently large number of neurons can approximate any smooth function under a reasonable set of assumptions (e.g., Cybenko 1989; Funahashi 1989; Hornik, Stinchcombe, and White 1989; Hornik 1991; Barron 1994). However, neural nets with one or two hidden layers—shallow neural nets—and a large number of neurons in each layer are often difficult to reliably fit (or train). In practice, neural nets with three or more hidden layers—deep neural nets (DNNs)—and a more limited number of neurons in each layer often perform better than shallow neural nets. DNNs can typically be trained more effectively than shallow neural nets, especially in the presence of a large number of predictors (as in our application).[37] Training a DNN for prediction is known as deep learning.

We consider two DNN architectures for predicting exchange rate changes, where our set of 70 predictors serves as the input layer for each network. The first, DNN3, contains three hidden layers with 16, eight, and four neurons, respectively, while the second, DNN4, contains four hidden layers with 16, eight, four, and two neurons, respectively.[38] DNN3 and DNN4 include 1,313 and 1,321 weights, respectively.

Training a DNN entails estimating the weights. We estimate the weights by minimizing an objective function based on mean squared prediction error for the training sample augmented by an ℓ_1 penalty term to better guard against overfitting. Computationally efficient algorithms based on stochastic gradient descent (SGD) are available for estimating the DNN weights; we use the recently developed Adam SGD algorithm (Kingma and Ba 2015). To implement the algorithm, we need to specify a number of hyperparameters, such as the dropout rate (Hinton et al. 2012; Srivastava et al. 2014), number of epochs, and batch size. Section A2 of the Internet Appendix provides details for the values that we use for the hyperparameters, which are set with an eye toward mitigating overfitting in our high-dimensional and noisy data environment.[39]

[37]See Rolnick and Tegmark (2018) and references therein for theoretical explanations of the advantages of deep over shallow neural nets.

[38]Our specification of 16 neurons in the first hidden layer is a compromise between two popular rules of thumb, namely, half or the square root of the number of predictors.

[39]Although it may be advantageous to use a validation sample to select the DNN architecture and hyperparameter values, this quickly becomes computationally impractical (e.g., Gu, Kelly, and Xiu 2020), especially since we train the DNNs recursively as more data become available.

6.2 Empirical Results

We generate the DNN forecasts analogously to the ENet forecasts. We first train the DNN using data through 1994:12 and use the fitted DNN to compute a forecast for each of the available exchange rates for 1995:01. Next, we train the DNN using data through 1995:01 and use the fitted DNN to compute a forecast for 1995:02. We continue in this manner through the end of the out-of-sample period.

Table 6 reports volatility ratios and R^2_{OS} statistics for the DNN3 and DNN4 forecasts. According to the third and fifth columns, the DNN3 forecasts are more volatile than the DNN4 forecasts. Nevertheless, both forecasts outperform the naïve no-change benchmark in terms of MSFE for the majority of countries. From the R^2_{OS} statistics in the fourth column of Table 6, we see that the DNN3 forecast outperforms the no-change benchmark for ten of the 14 countries—including eight of the G10 countries—and most of the improvements are significant. The R^2_{OS} statistics are sizable in many cases, with values of 2.25% or more for seven countries. For the entire group of countries, the R^2_{OS} statistic is 1.44% (significant at the 1% level). However, the negative R^2_{OS} statistics are sizable in magnitude for Canada, Germany, and the Netherlands (−2.24%, −4.89%, and −6.28%, respectively), so that the performance of the DNN3 forecast is somewhat erratic.

According to the last column of Table 6, the DNN4 forecast also outperforms the no-change benchmark for ten of the 14 countries. Among the G10 countries, it only fails to outperform the benchmark for Japan. A number of the R^2_{OS} statistics are again significant and sizable in magnitude. Overall, compared to those for the DNN3 forecast, the results for the DNN4 forecast are more stable, and the R^2_{OS} statistics are typically smaller in magnitude for the latter vis-à-vis the former. For the 14 countries taken together, the R^2_{OS} statistic is 1.00% for the DNN4 forecast (significant at the 1% level).

Comparing the R^2_{OS} statistics in the fourth and sixth columns of Table 6 with those in the last column of Table 3, the ENet-ERIC forecast appears to perform better overall than the DNN forecasts. Although the DNN forecasts produce a larger R^2_{OS} statistic in some instances, they also underperform the no-change benchmark for some countries by a sizable margin. These results suggest that the degree of nonlinearity in the predictive relationships is generally not strong enough in our application to outweigh the greater susceptibility to overfitting inherent in the DNNs, given the need to estimate a very large number of weights. Section 6.3 presents further evidence in support of this interpretation.[40]

6.3 Peering into the Black Box

The partial dependence plot (PDP, Friedman 2001) can be used to examine the relationship between the expected value of the target variable and a given predictor for any fitted model, including black box models like DNNs. Let $\hat{f}(\boldsymbol{x})$ denote the prediction function for a fitted model, such as a DNN. The PDP for x_q is defined as

$$\begin{aligned} f_q(x_q) &= E_{\boldsymbol{x}_{C(q)}}\left[\hat{f}\big(x_q, \boldsymbol{x}_{C(q)}\big)\right] \\ &= \int_{\boldsymbol{x}_{C(q)}} \hat{f}\big(x_q, \boldsymbol{x}_{C(q)}\big) p_{C(q)}\big(\boldsymbol{x}_{C(q)}\big)\, d\boldsymbol{x}_{C(q)}, \end{aligned} \tag{6.4}$$

where $\boldsymbol{x}_{C(q)} = \boldsymbol{x} \setminus x_q$, and

$$p_{C(q)}\big(\boldsymbol{x}_{C(q)}\big) = \int_{x_q} p(\boldsymbol{x})\, dx_q \tag{6.5}$$

is the joint marginal probability density for $\boldsymbol{x}_{C(q)}$. Equation (6.4) is typically estimated via Monte Carlo integration using the training data, which in our panel data setting is given by

$$\bar{f}_q(x_q) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{f}\big(x_q, \boldsymbol{x}_{i,t,C(q)}\big). \tag{6.6}$$

It is common to plot the PDP for the training sample values of the predictor of interest (i.e., $x_{i,t,q}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$) or specific quantiles for the training sample.

[40] We also constructed Smart-Opt portfolios based on the DNN3 and DNN4 forecasts. The portfolios typically outperform the Basic-Opt portfolio, but they do not perform as well as the Smart-Opt portfolio based on the ENet-ERIC forecast; see Tables A11 and A12 of the Internet Appendix.

Greenwell, Boehmke, and McCarthy (2018) propose a partial dependence-based measure of predictor importance. In our panel framework, it is given by

$$\mathcal{I}(x_q) = \left\{ \frac{1}{NT-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\bar{f}_q(x_{i,t,q}) - \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \bar{f}_q(x_{i,t,q}) \right]^2 \right\}^{0.5}. \tag{6.7}$$

Intuitively, Equation (6.7) uses the sample standard deviation of the PDP to measure its "flatness." If the expected value of the target does not change as the predictor changes, then the PDP is flat, so that the variable importance measure is zero. As the variability of the PDP increases, the importance measure in Equation (6.7) likewise increases.

The leftmost panel of Figure 11 shows the Greenwell, Boehmke, and McCarthy (2018) variable importance measure for each predictor for the linear panel predictive regression, Equation (3.6), estimated via the ENet-ERIC; the center and rightmost panels show the measures for the fitted DNN3 and DNN4. The fitted models correspond to the final recursive estimates (which use the most data). The importance measures are exactly zero in the leftmost panel for the predictors that are not selected by the ERIC-ENet. Many of the same predictors evince spikes across the three panels of Figure 11, so that there is relatively strong commonality in the importance of the predictors across the fitted models (especially BILL.GVOL, INF.GVOL, IV.GCOR, DP, PE.GILL, PE, and SRET.MPU). The predictors that are not selected by the ENet-ERIC are often of limited importance in the fitted DNNs.

Figure 12 depicts PDPs for the fitted models. For the linear model, the plots have a constant slope (by construction); the plots are horizontal lines at zero for the predictors that are not selected by the ENet-ERIC. For the predictors selected by ENet-ERIC, the PDPs for the fitted DNNs frequently lie close to those for the fitted linear model, so that the predictors exhibit mild nonlinearities in the DNNs (e.g., DP, INF.GVOL, UN.GCOR, PE.EPU, PE.GILL, SRET.MPU, and IV.GCOR). When the plots exhibit more obvious nonlinearities for the fitted DNNs, the nonlinearities often appear quantitatively modest.[41]

[41] Note the different scales for the vertical axes in Figure 12.

Figures 11 and 12 allow us to peer into the black box of the fitted DNN models to investigate the importance of individual predictors and the strength of nonlinearities in the predictive relationships. The figures indicate that many of the same predictors appear important in the fitted linear and DDN models and that the nonlinearities in the latter are relatively weak. The fitted models thus appear fairly "close" to one another, with the DNNs including a plethora of additional—and relatively weak—predictive relationships that likely result from the difficulty in training a model with a plethora of weights in a noisy data environment. Overall, in our context, the nonlinearities in the fitted DNNs do not appear strong enough to reliably improve out-of-sample performance vis-à-vis the linear ENet-ERIC forecast.

7 Conclusion

Short-horizon exchange rate prediction has posed an enduring challenge to researchers in international finance. In this paper, we make considerable progress in resolving the Meese and Rogoff (1983) no-predictability puzzle by showing that monthly US dollar exchange rate forecasts can significantly outperform the naïve no-change benchmark forecast over a lengthy out-of-sample period for a group of 14 developed countries. There are three key elements in our forecasting approach. First, we consider a rich information set, which includes ten country characteristics and six global variables; after interacting the country characteristics with the global variables, we have 70 predictors. It is important to consider a large number of potential predictors, rather than only a few fundamentals, as we cannot know ex ante which predictors are the most relevant. Second, we employ the ENet, a machine learning device for implementing penalized regression, to guard against overfitting the panel predictive regression that we use to generate the forecasts. Our high-dimensional setting with 70 predictors and the substantial unpredictable component (i.e., noise) in monthly change rate changes make it vital to guard against overfitting; indeed, the ENet-ERIC forecast—which imposes the strongest degree of shrinkage—performs the best overall. Third, we impose the pooling restriction when estimating the panel predictive regression, which substantially reduces the number of parameters that we need to estimate, thereby further alleviating overfitting.

In addition to improving out-of-sample prediction in terms of MSFE, we show that the ENet-ERIC forecast provides substantive economic value to a US investor. Specifically, the performance of an optimal portfolio for a mean-variance investor who allocates across foreign currencies improves markedly when the investor utilizes the ENet-ERIC forecast of the exchange rate change when predicting the currency excess return. The ENet-ERIC forecast is especially valuable to the investor during and after the global financial crisis. During the worst phase of the crisis in late 2008, the ENet-ERIC forecast generates sizable improvements in portfolio performance by anticipating a sharp devaluation in many foreign currencies, which likely results from a decrease in global risk tolerance around the crisis and the US dollar's safe-haven role.

We use another machine learning technique, deep learning, to allow for more complex predictive relationships when forecasting exchange rates. Although the DNN forecasts outperform the no-change benchmark for most countries, they generally do not perform as well as the ENet-ERIC forecast. Variable importance measures and partial dependence plots reveal that the nonlinearities in the fitted DNNs are relatively modest. The nonlinearities thus appear too weak to improve out-of-sample performance compared to the ENet-ERIC forecast, given the challenges in training forecasting models with a plethora of parameters (like DNNs) in a noisy data environment.

Our fresh evidence of out-of-sample exchange range predictability raises fundamental issues in international finance. What are the theoretical underpinnings of exchange rate predictability? To what extent does predictability reflect rational time-varying risk premia or mispricing in the foreign exchange market? Do arbitrage frictions play a significant role, even though transactions costs are relatively small in major currency markets? In light of our new evidence, we view these questions as important topics for future research.

References

Adrian, T. and P. Xie (2020). The Non-U.S. Bank Demand for U.S. Dollar Assets. CEPR Working Paper.

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In *Proceedings of the 2nd International Symposium on Information Theory*. Ed. by B. N. Petrov and F. Csaki. Budapest: Akadémiai Kiadó, pp. 267–281.

Ang, A. and J. Chen (2011). Yield Curve Predictors of Foreign Exchange Returns. Working Paper.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring Economic Policy Uncertainty. *Quarterly Journal of Economics* 131:4, 1593–1636.

Bakshi, G. and G. Panayotov (2013). Predictability of Currency Carry Trades and Asset Pricing Implications. *Journal of Financial Economics* 110:1, 139–163.

Bao, Y. (2009). Estimation Risk-Adjusted Sharpe Ratio and Fund Performance Ranking under a General Return Distribution. *Journal of Financial Econometrics* 7:2, 152–173.

Barron, A. R. (1994). Approximation and Estimation Bounds for Artificial Neural Networks. *Machine Learning* 14:1, 115–133.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). Simultaneous Analysis of Lasso and Dantzig Selector. *Annals of Statistics* 37:4, 1705–1732.

Bilson, J. F. O. (1981). The 'Speculative Efficiency' Hypothesis. *Journal of Business* 54:3, 435–451.

Brunnermeier, M. K., S. Nagel, and L. H. Pedersen (2009). Carry Trades and Currency Crashes. In *NBER Macroeconomics Annual 2008*. Ed. by D. Acemoglu, K. Rogoff, and M. Woodford. Vol. 23. Chicago: University of Chicago Press, pp. 313–347.

Burnside, C. (2018). Exchange Rates, Interest Parity, and the Carry Trade. Working Paper.

Burnside, C., M. Eichenbaum, I. Kleshchelski, and S. Rebelo (2011). Do Peso Problems Explain the Returns to the Carry Trade? *Review of Financial Studies* 24:3, 853–891.

Burnside, C., M. Eichenbaum, and S. Rebelo (2011). Carry Trade and Momentum in Currency Markets. *Annual Review of Finance and Economics* 3:1, 511–535.

Caldara, D. and M. Iacoviello (2018). Measuring Geopolitical Risk. Federal Reserve Board of Governors International Finance Discussion Paper No. 1222.

Campbell, J. Y. and S. B. Thompson (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies* 21:4, 1509–1531.

Cenedese, G., R. Payne, L. Sarno, and G. Valente (2016). What Do Stock Markets Tell Us About Exchange Rates? *Review of Finance* 20:3, 1045–1080.

Chen, Y.-C. and K. P. Tsang (2013). What Does the Yield Curve Tell Us About Exchange Rate Predictability? *Review of Economics and Statistics* 95:1, 185–205.

Christiano, L. J. and T. J. Fitzgerald (2003). The Band Pass Filter. *International Economic Review* 44:2, 435–465.

Clark, T. E. and M. W. McCracken (2001). Tests of Equal Forecast Accuracy and Ecompassing for Nested Models. *Journal of Econometrics* 105:1, 85–110.

Clark, T. E. and K. D. West (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics* 138:1, 291–311.

Cybenko, G. (1989). Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems* 2:4, 303–314.

Dahlquist, M. and H. Hasseltoft (2020). Economic Momentum and Currency Returns. *Journal of Financial Economics* 136:1, 152–167.

Daniel, K., R. J. Hodrick, and Z. Lu (2017). The Carry Trade: Risks and Drawdowns. *Critical Finance Review* 6:2, 211–262.

Della Corte, P., L. Sarno, and I. Tsiakas (2009). An Economic Evaluation of Empirical Exchange Rate Models. *Review of Financial Studies* 22:9, 3491–3530.

Diebold, F. X. and R. S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13:3, 253–263.

Dobrynskaya, V. (2014). Downside Market Risk of Carry Trades. *Review of Finance* 18:5, 1885–1913.

Engel, C. and K. D. West (2005). Exchange Rates and Fundamentals. *Journal of Political Economy* 113:3, 485–517.

Engel, C. and S. P. Y. Wu (2019). Liquidity and Exchange Rates: An Empirical Investigation. NBER Working Paper No. 25397.

Fama, E. F. (1984). Forward and Spot Exchange Rates. *Journal of Monetary Economics* 14:3, 319–338.

Fan, Y. and C. Y. Tang (2013). Tuning Parameter Selection in High Dimensional Penalized Likelihood. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 75:3, 531–552.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the Factor Zoo: A Test of New Factors. *Journal of Finance* 75:3, 1327–1370.

Filippou, I. and M. P. Taylor (2017). Common Macro Factors and Currency Premia. *Journal of Financial and Quantitative Analysis* 52:4, 1731–1763.

Flynn, C. J., C. M. Hurvich, and J. S. Simonoff (2013). Efficiency for Regularization Parameter Selection in Penalized Likelihood Estimation of Misspecified Models. *Journal of the American Statistical Association* 108:503, 1031–1043.

Frenkel, J. A. (1976). A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence. *Scandinavian Journal of Economics* 78:2, 200–224.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting Characteristics Nonparametrically. *Review of Financial Studies* 33: 2326–2377.

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* 29:5, 1189–1232.

Froot, K. A. and R. H. Thaler (1990). Anomalies: Foreign Exchange. *Journal of Economic Perspectives* 4:3, 179–192.

Funahashi, K.-I. (1989). On the Approximate Realization of Continuous Mappings by Neural Networks. *Neural Networks* 2:3, 183–192.

Goyal, A. and A. Saretto (2009). Cross-Section of Option Returns and Volatility. *Journal of Financial Economics* 94:2, 310–326.

Goyal, A. and I. Welch (2003). Predicting the Equity Premium with Dividend Ratios. *Management Science* 49:5, 639–654.

Goyal, A. and I. Welch (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* 21:4, 1455–1508.

Greenwell, B. M., B. C. Boehmke, and A. J. McCarthy (2018). A Simple and Effective Model-Based Variable Importance Measure. arXiv:1805.04755v1.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical Asset Pricing via Machine Learning. *Review of Financial Studies* 33:5, 2223–2273.

Han, Y., A. He, D. E. Rapach, and G. Zhou (2020). Firm Characteristics and Expected Stock Returns. Working Paper.

Hansen, L. P. and R. J. Hodrick (1980). Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis. *Journal of Political Economy* 88:5, 829–853.

Hastie, T. and J. Qian (2016). Glmnet Vignette. Working Paper.

Hastie, T., R. Tibshirani, and M. Wainwright (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Boca Raton, FL: CRC Press.

Hau, H. and H. Rey (2006). Exchange Rates, Equity Prices, and Capital Flows. *Review of Financial Studies* 19:1, 273–317.

Hinton, G. E., N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov (2012). Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors. Working paper (arXiv:1207.0580).

Hjalmarsson, E. (2010). Predicting Global Stock Returns. *Journal of Financial and Quantitative Analysis* 45:1, 49–80.

Hodrick, R. J. and E. C. Prescott (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit, and Banking* 19:1, 1–16.

Hoerl, A. E. and R. W. Kennard (1970). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* 12:1, 69–82.

Hornik, K. (1991). Approximation Capabilities of Multilayer Feedforward Networks. *Neural Networks* 4:2, 251–257.

Hornik, K., M. Stinchcombe, and H. White (1989). Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks* 2:5, 359–366.

Hui, F. K. C., D. I. Warton, and S. D. Foster (2015). Tuning Parameter Selection for the Adaptive Lasso Using ERIC. *Journal of the American Statistical Society* 110:509, 262–269.

Hurvich, C. M. and C.-L. Tsai (1989). Regression and Time Series Model Selection in Small Samples. *Biometrika* 76:2, 297–307.

Jiang, Z., A. Krishnamurthy, and H. Lustig (2019). Foreign Safe Asset Demand and the Dollar Exchange Rate. NBER Working Paper No. 24439.

Jordà, Ò. and A. M. Taylor (2012). The Carry Trade and Fundamentals: Nothing to Fear but FEER Itself. *Journal of International Economics* 88:1, 74–90.

Jurek, J. W. (2014). Crash-Neutral Currency Carry Trades. *Journal of Financial Economics* 113:3, 325–347.

Kingma, D. P. and J. Ba (2015). *Adam: A Method for Stochastic Optimization.* Third Annual International Conference on Learning Representations. San Diego.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the Cross Section. *Journal of Financial Economics* 135:2, 271–292.

Kremens, L. and I. Martin (2019). The Quanto Theory of Exchange Rates. *American Economic Review* 109:3, 810–843.

Lettau, M., M. Maggiori, and M. Weber (2014). Conditional Risk Premia in Currency Markets and Other Asset Classes. *Journal of Financial Economics* 114:2, 197–225.

Lilley, A., M. Maggiori, B. Neiman, and J. Schreger (forthcoming). Exchange Rate Reconnect. *Review of Economics and Statistics*.

Lustig, H., N. Roussanov, and A. Verdelhan (2011). Common Risk Factors in Currency Markets. *Review of Financial Studies* 24:11, 3731–3777.

Lyons, R. K. (2001). *The Microstructure Approach to Exchange Rates*. Cambridge, MA: MIT Press.

Mark, N. C. (1995). Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability. *American Economic Review* 85:1, 201–218.

McCracken, M. W. (2007). Asymptotics for Out of Sample Tests of Granger Causality. *Journal of Econometrics* 140:2, 719–752.

Meese, R. A. and K. Rogoff (1983). Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample? *Journal of International Economics* 14:1–2, 3–24.

Meinshausen, N. and B. Yu (2009). Lasso-Type Recovery of Sparse Representations for High-Dimensional Data. *Annals of Statistics* 37:1, 246–270.

Melvin, M. and D. Shand (2017). When Carry Goes Bad: The Magnitude, Causes, and Duration of Currency Carry Unwinds. *Financial Analysts Journal* 73:1, 121–144.

Melvin, M. and M. P. Taylor (2009). The Crisis in the Foreign Exchange Market. *Journal of International Money and Finance* 28:8, 1317–1330.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012a). Carry Trades and Global Foreign Exchange Volatility. *Journal of Finance* 67:2, 681–718.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012b). Currency Momentum Strategies. *Journal of Financial Economics* 106:3, 660–684.

Molodtsova, T. and D. H. Papell (2009). Out-of-Sample Exchange Rate Predictability with Taylor Rule Fundamentals. *Journal of International Economics* 77:2, 167–180.

Mueller, P., A. Stathopoulos, and A. Vedolin (2017). International Correlation Risk. *Journal of Financial Economics* 126:2, 270–299.

Mussa, M. (1976). The Exchange Rate, the Balance of Payments and Monetary and Fiscal Policy under a Regime of Controlled Floating. *Scandinavian Journal of Economics* 78:2, 229–248.

Neely, C. J. and P. A. Weller (2013). Lessons from the Evolution of Foreign Exchange Trading Strategies. *Journal of Banking and Finance* 37:10, 3783–3798.

Neely, C. J., P. A. Weller, and J. M. Ulrich (2009). The Adaptive Markets Hypothesis: Evidence from the Foreign Exchange Market. *Journal of Financial and Quantitative Analysis* 44:2, 467–488.

Rapach, D. E., J. K. Strauss, and G. Zhou (2013). International Stock Return Predictability: What Is the Role of the United States? *Journal of Finance* 68:4, 1633–1662.

Rapach, D. E. and G. Zhou (2013). Forecasting Stock Returns. In *Handbook of Economic Forecasting*. Ed. by G. Elliott and A. Timmermann. Vol. 2A. Amsterdam: Elsevier, pp. 328–383.

Rolnick, D. and M. Tegmark (2018). *The Power of Deeper Networks for Expressing Natural Functions*. Sixth Annual International Conference on Learning Representations. Vancouver.

Rossi, B. (2013). Exchange Rate Predictability. *Journal of Economic Literature* 51:4, 1063–1119.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics* 6:2, 461–464.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15:56, 1929–1958.

Taddy, M. (2017). One-Step Estimator Paths for Concave Regularization. *Journal of Computational and Graphical Statistics* 26:3, 525–536.

Taylor, J. B. (1993). Discretion Versus Policy Rules in Practice. *Carnegie-Rochester Conference Series on Public Policy* 39:1, 195–214.

Taylor, M. P. (1995). The Economics of Exchange Rates. *Journal of Economic Literature* 33:1, 13–47.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)* 58:1, 267–288.

Wang, H., B. Li, and C. Leng (2009). Shrinkage Tuning Parameter Selection with a Diverging Number of Parameters. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 71:3, 671–683.

West, K. D. (1996). Asymptotic Inference about Predictive Ability. *Econometrica* 64:5, 1067–1084.

Zhang, C.-H. and J. Huang (2008). The Sparsity and Bias of the Lasso Selection in High-Dimensional Linear Regression. *Annals of Statistics* 36:4, 1567–1594.

Zou, H. and T. Hastie (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67:2, 301–320.

Zou, H., T. Hastie, and R. Tibshirani (2007). On the 'Degrees of Freedom' of the Lasso. *Annals of Statistics* 35:5, 2173–2192.

Table 1: Summary statistics for exchange rate changes

(1)	(2)	(3)	(4)	(5)
Country	Sample period	Annualized mean	Annualized volatility	Autocorrelation
United Kingdom	1985:01–2019:03	0.16%	10.05%	0.07
Switzerland	1985:01–2019:03	−2.16%	11.29%	−0.01
Japan	1985:01–2019:03	−1.78%	11.02%	0.04
Canada	1985:01–2019:03	0.30%	7.39%	−0.04
Australia	1985:01–2019:03	1.12%	11.80%	0.05
New Zealand	1985:01–2019:03	−0.31%	12.18%	−0.03
Sweden	1985:01–2019:03	0.71%	11.16%	0.11
Norway	1985:01–2019:03	0.43%	10.92%	0.03
Denmark	1985:01–2019:03	−1.00%	10.44%	0.03
Euro area	1999:02–2019:03	0.54%	9.80%	0.04
Germany	1985:01–1998:12	−3.87%	11.64%	0.03
Italy	1985:01–1998:12	−0.48%	11.54%	0.10
France	1985:01–1998:12	−3.27%	11.18%	0.02
Netherlands	1985:01–1998:12	−3.89%	11.59%	0.04

The table reports summary statistics for monthly exchange rate changes measured against the US dollar. The country-i exchange rate change is defined as $(S_{i,t}/S_{i,t-1})-1$, where $S_{i,t}$ is the month-t spot exchange rate for country i (number of country-i currency units per US dollar). The annualized mean (volatility) in the third (fourth) column is the monthly mean (standard deviation) multiplied by 12 $\left(\sqrt{12}\right)$.

Table 2: Volatility ratios

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Country	Obs.	OLS	ENet-BIC	ENet-MBIC	ENet-GIC	ENet-ERIC
United Kingdom	291	0.54	0.18	0.16	0.18	0.14
Switzerland	291	0.39	0.14	0.12	0.14	0.10
Japan	291	0.31	0.12	0.09	0.11	0.08
Canada	291	0.51	0.21	0.20	0.21	0.16
Australia	291	0.36	0.16	0.14	0.15	0.12
New Zealand	291	0.40	0.15	0.14	0.15	0.12
Sweden	291	0.38	0.15	0.14	0.14	0.12
Norway	291	0.41	0.18	0.16	0.18	0.14
Denmark	291	0.41	0.15	0.13	0.15	0.12
Euro area	230	0.49	0.24	0.22	0.24	0.20
Germany	48	0.33	0.04	0.03	0.03	0.03
Italy	48	0.31	0.08	0.04	0.07	0.04
France	48	0.49	0.08	0.03	0.04	0.03
Netherlands	48	0.39	0.06	0.05	0.06	0.05
All	3,041	0.41	0.17	0.15	0.16	0.13

The table reports volatility ratios for monthly out-of-sample forecasts of exchange rate changes based on a panel predictive regression estimated via ordinary least squares (OLS) and the elastic net (ENet). The regularization parameter for the elastic net is tuned via the Bayesian information criterion (BIC), modified BIC (MBIC), generalized information criterion (GIC), or extended regularization information criterion (ERIC). The forecast is based on a panel predictive regression with 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables. The volatility ratio is the standard deviation for the forecast divided by the standard deviation for the actual exchange rate change. The second column reports the number of out-of-sample observations.

Table 3: R^2_{OS} statistics (in percent)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Country	Obs.	OLS	ENet-BIC	ENet-MBIC	ENet-GIC	ENet-ERIC
United Kingdom	291	-10.83	4.79^{**}	4.09^{*}	4.85^{**}	5.32^{**}
Switzerland	291	-11.78	1.55^{*}	0.76	1.24^{*}	1.95^{*}
Japan	291	-10.53	-0.34	-0.70	-0.92	0.31
Canada	291	-18.76	-1.19	-1.07	-1.14	0.06
Australia	291	-3.57	1.30^{**}	0.69^{*}	1.02^{**}	1.34^{*}
New Zealand	291	-5.41	3.26^{**}	2.41^{*}	3.09^{**}	2.61^{*}
Sweden	291	-2.60	2.15^{**}	2.95^{**}	2.30^{**}	2.83^{**}
Norway	291	-1.88	2.38^{**}	2.14^{**}	2.16^{**}	3.03^{**}
Denmark	291	-6.22	2.36^{***}	1.42^{**}	2.22^{***}	1.99^{***}
Euro area	230	-10.76	1.87^{**}	1.42^{*}	1.71^{**}	2.15^{*}
Germany	48	-11.59	-0.35	0.18	0.25	0.25
Italy	48	-12.53	-0.50	-0.16	-0.43	-0.33
France	48	-36.64	-1.36	-0.03	-0.45	0.39
Netherlands	48	-21.81	-1.69	1.29	1.36	1.05
All	3,041	-8.03	1.72^{***}	1.38^{***}	1.60^{***}	2.04^{***}

The table reports out-of-sample R^2 (R^2_{OS}) statistics for monthly out-of-sample forecasts of exchange rate changes based on a panel predictive regression estimated via ordinary least squares (OLS) and the elastic net (ENet). The regularization parameter for the elastic net is tuned via the Bayesian information criterion (BIC), modified BIC (MBIC), generalized information criterion (GIC), or extended regularization information criterion (ERIC). The R^2_{OS} statistic is the percent reduction in mean squared forecast error (MSFE) for a competing forecast vis-á-vis the no-change benchmark forecast. The competing forecast is based on a panel predictive regression with 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables. For a positive R^2_{OS} statistic, we use the Clark and West (2007) MSFE-adj statistic to test the null hypothesis that the benchmark MSFE is less than or equal to the competing MSFE against the alternative hypothesis that the benchmark MSFE is greater than the competing MSFE; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The second column reports the number of out-of-sample observations.

Table 4: Portfolio performance

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Basic-Opt portfolio			Smart-Opt portfolio		
Transaction costs	Annualized mean	Annualized volatility	Annualized Sharpe ratio	Annualized mean	Annualized volatility	Annualized Sharpe ratio
Panel A: Full sample (1995:01 to 2019:03)						
Mid-quotes	7.47%	10.22%	0.73***	12.94%	12.80%	1.01***
25%	5.69%	10.13%	0.56***	10.54%	12.81%	0.82***
50%	4.92%	10.11%	0.49**	9.71%	12.80%	0.76***
75%	4.16%	10.11%	0.41**	8.89%	12.80%	0.69***
Bid-ask spread	3.39%	10.10%	0.34*	8.06%	12.80%	0.63***
Panel B: Pre-crisis subsample (1995:01 to 2008:08)						
Mid-quotes	11.88%	10.56%	1.13***	13.82%	12.18%	1.13***
25%	9.80%	10.58%	0.93***	10.99%	12.47%	0.88***
50%	8.81%	10.59%	0.83***	9.93%	12.47%	0.80***
75%	7.82%	10.59%	0.74***	8.88%	12.47%	0.71***
Bid-ask spread	6.83%	10.59%	0.64**	7.82%	12.48%	0.63**
Panel C: Post-crisis subsample (2008:09 to 2019:03)						
Mid-quotes	1.76%	9.55%	0.18	11.81%	13.60%	0.87***
25%	0.39%	9.32%	0.04	9.96%	13.28%	0.75**
50%	−0.09%	9.32%	−0.01	9.43%	13.27%	0.71**
75%	−0.57%	9.31%	−0.06	8.90%	13.26%	0.67**
Bid-ask spread	−1.04%	9.31%	−0.11	8.37%	13.25%	0.63**

The table reports annualized summary statistics for the portfolio excess return for a US investor with mean-variance preferences and a relative risk aversion coefficient of five who invests in 14 foreign currencies. The Basic-Opt (Smart-Opt) portfolio assumes that the investor uses the no-change (elastic net) exchange rate forecast. The elastic net forecast is based on elastic net estimation of a panel predictive regression with 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables, where the regularization parameter is tuned via the extended regularization information criterion. Mid-quotes ignore transaction costs. We account for transaction costs using bid-ask spreads from Datastream. Because the full bid-ask spreads likely overstate transaction costs, we compute results assuming 25%, 50%, and 75% of the full bid-ask spreads. We test the significance of the Sharpe ratios using t-statistics based on Bao (2009) standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Alphas and currency factor exposures

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Basic-Opt portfolio				Smart-Opt portfolio			
Sample	Ann. alpha	MKT_{FX}	HML_{FX}	$\bar{R}^2$	Ann. alpha	MKT_{FX}	HML_{FX}	$\bar{R}^2$
1995:01 to 2019:03	3.82%	−0.001	0.75	52.19%	10.83%	−0.52	0.48	16.90%
	[2.29]**	[−0.01]	[9.96]***		[3.18]***	[−2.80]***	[2.75]***	
1995:01 to 2008:08	5.17%	0.17	0.96	65.20%	8.54%	−0.42	0.88	47.77%
	[5.27]***	[1.78]*	[12.11]***		[3.13]***	[−2.42]**	[8.86]***	
2008:09 to 2019:03	0.25%	−0.04	0.59	41.93%	11.41%	−0.31	0.05	1.83%
	[0.11]	[−0.29]	[4.49]***		[2.41]**	[−1.40]	[0.19]	

The table reports Lustig, Roussanov, and Verdelhan (2011) currency factor model estimation results for the portfolio excess return for a US investor with mean-variance preferences and a relative risk aversion coefficient of five who invests in 14 foreign currencies. The Basic-Opt (Smart-Opt) portfolio assumes that the investor uses the no-change (elastic net) exchange rate forecast. The elastic net forecast is based on elastic net estimation of a panel predictive regression with 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables, where the regularization parameter is tuned via the extended regularization information criterion. MKT_{FX} (HML_{FX}) is the dollar (carry trade) risk factor. Brackets report t-statistics based on robust standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Out-of-sample results for deep neural network forecasts

(1)	(2)	(3)	(4)	(5)	(6)
		DNN3		DNN4	
Country	Obs.	Volatility ratio	R^2_{OS} (%)	Volatility ratio	R^2_{OS} (%)
United Kingdom	291	0.30	0.87*	0.13	2.26**
Switzerland	291	0.21	2.25**	0.10	0.21
Japan	291	0.18	0.21*	0.09	−1.38
Canada	291	0.29	−2.24	0.13	0.92*
Australia	291	0.21	1.09***	0.09	0.70
New Zealand	291	0.19	2.54**	0.09	2.15***
Sweden	291	0.19	3.35***	0.09	2.51***
Norway	291	0.22	2.70***	0.11	1.25*
Denmark	291	0.21	2.90***	0.11	1.35**
Euro area	230	0.27	−0.14	0.12	1.90**
Germany	48	0.13	−4.89	0.05	−4.41
Italy	48	0.22	5.35*	0.12	1.29
France	48	0.17	2.91	0.09	−0.14
Netherlands	48	0.12	−6.28	0.05	−3.89
All	3,041	0.22	1.44***	0.11	1.00***

The table reports out-of-sample results for monthly forecasts of exchange rate changes based on deep neural networks (DNNs). For each DNN, 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables serve as the input layer. DNN3 contains three hidden layers with 16, eight, and four neurons, respectively; DNN4 contains four hidden layers with 16, eight, four, and two neurons, respectively. The volatility ratio is the standard deviation for the forecast divided by the standard deviation for the actual exchange rate change. The out-of-sample R^2 (R^2_{OS}) statistic is the percent reduction in mean squared forecast error (MSFE) for a competing forecast vis-á-vis the no-change benchmark forecast. For a positive R^2_{OS} statistic, we use the Clark and West (2007) MSFE-adj statistic to test the null hypothesis that the benchmark MSFE is less than or equal to the competing MSFE against the alternative hypothesis that the benchmark MSFE is greater than the competing MSFE; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The second column reports the number of out-of-sample observations.

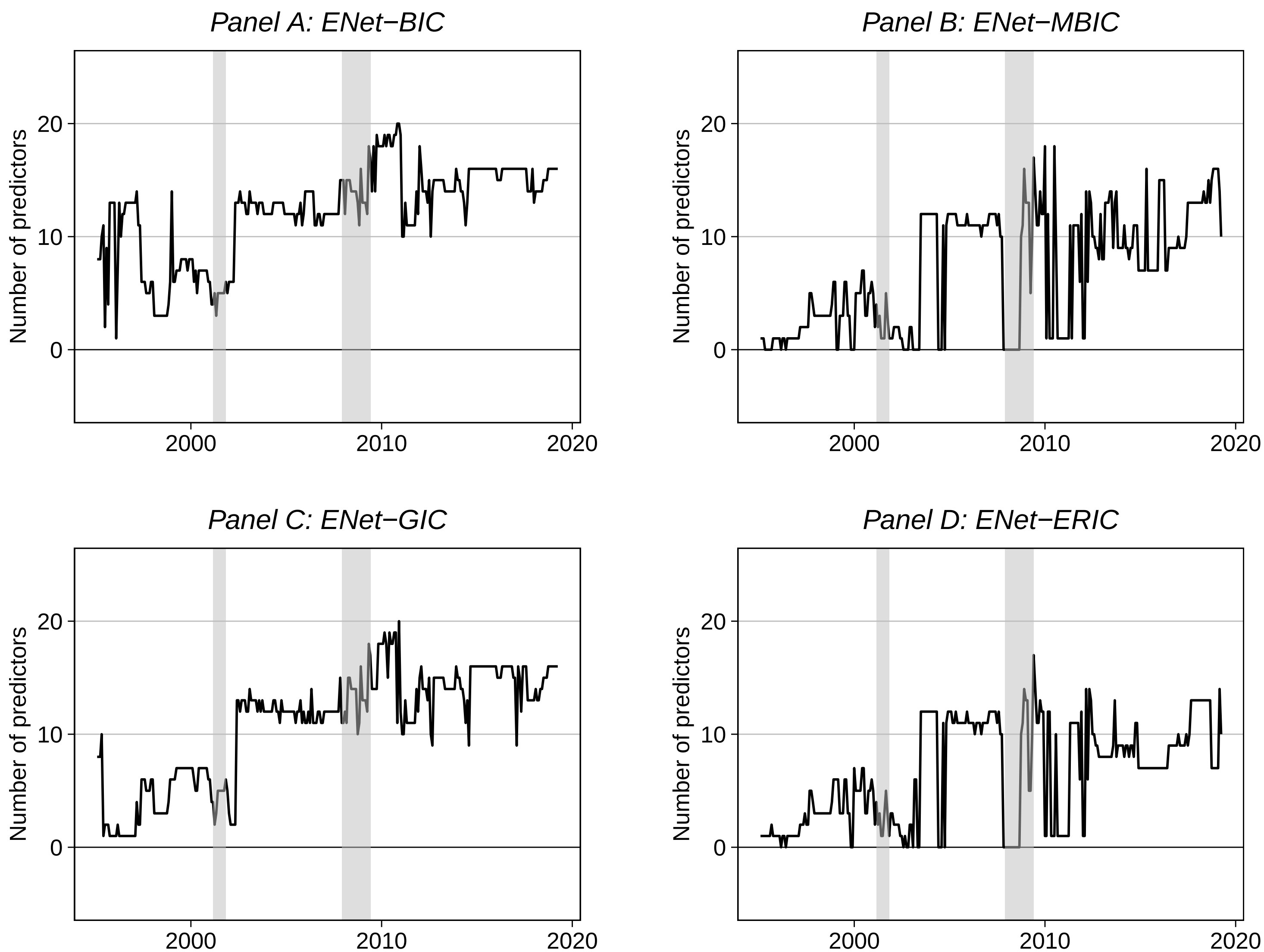


Figure 1: Number of predictors selected by the elastic net

The figure shows the number of predictors selected by the elastic net for recursive estimation of a panel predictive regression with 70 predictors. Results are shown for different criteria for tuning the regularization parameter. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

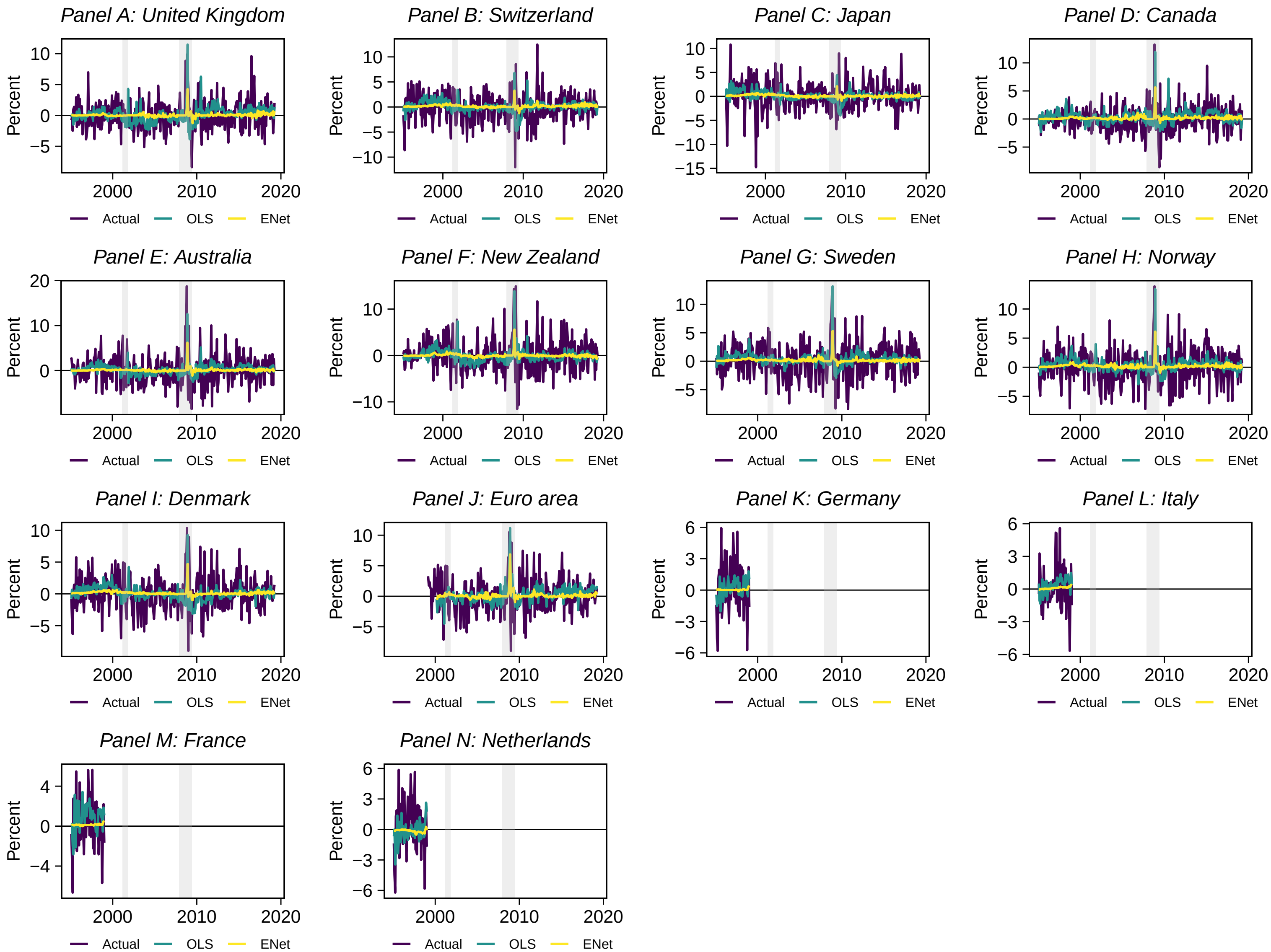


Figure 2: OLS and ENet-ERIC forecasts

Each panel shows monthly out-of-sample forecasts of exchange rate changes based on recursive ordinary least squares (OLS) and elastic net (ENet) estimation of a panel predictive regression with 70 predictors, where the regularization parameter for the elastic net is tuned via the extended regularization information criterion. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

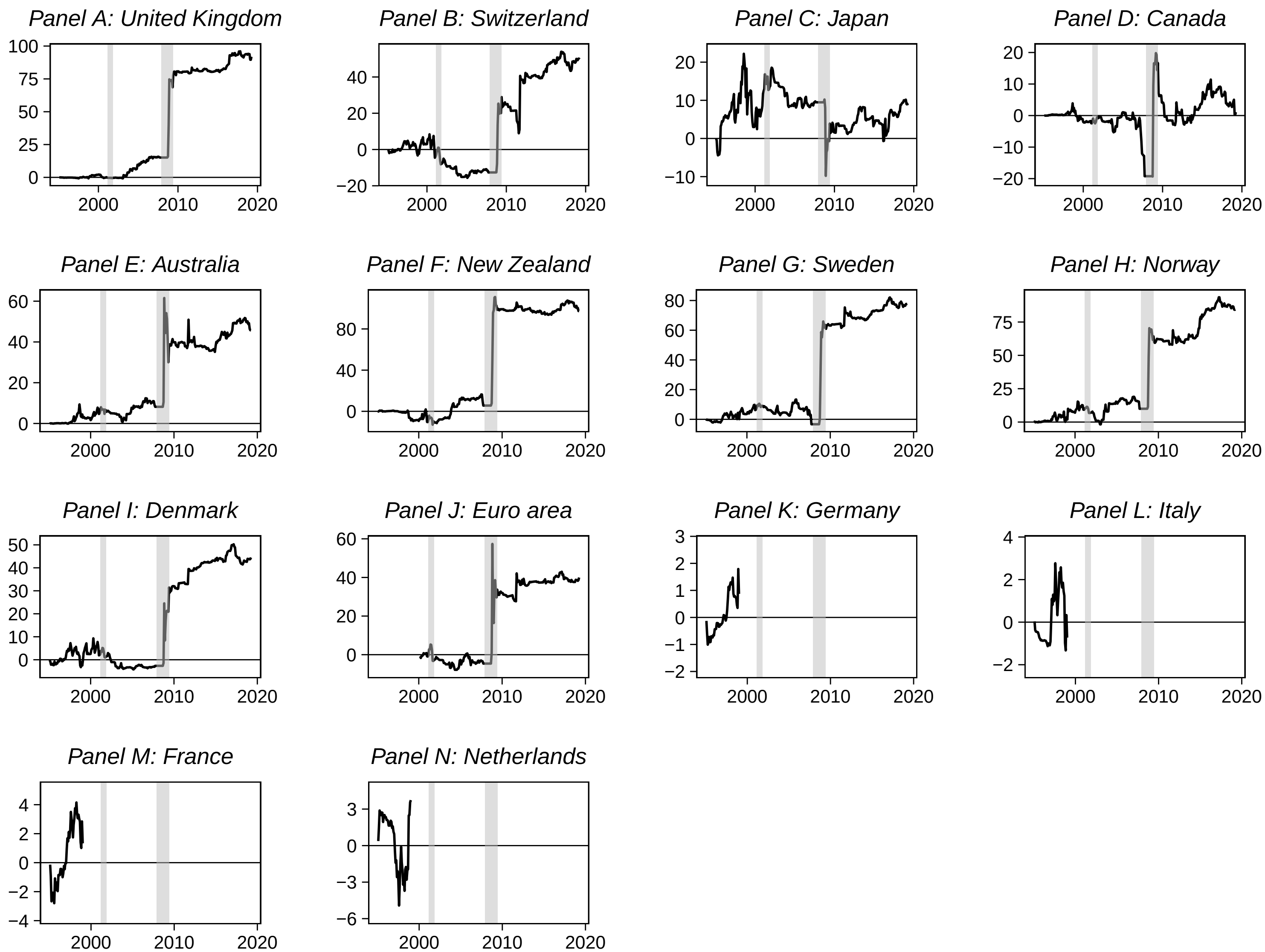


Figure 3: Cumulative difference in squared forecast errors

The figure shows the cumulative difference in squared forecast errors for the no-change benchmark forecast vis-à-vis the ENet-ERIC forecast. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

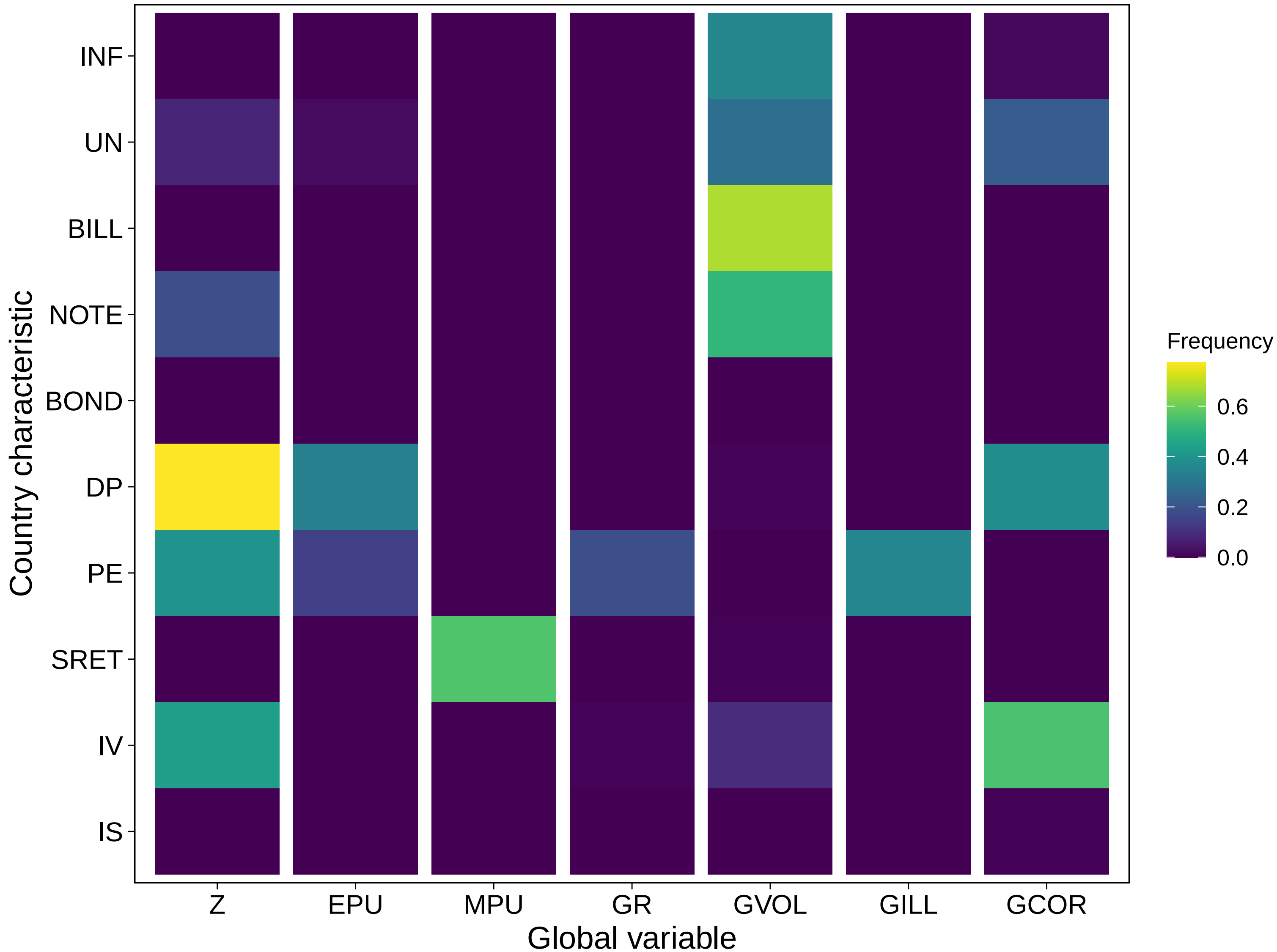


Figure 4: Heatmap for predictors selected by the elastic net

The figure shows the frequencies for predictors selected by the elastic net when the regularization parameter is tuned via the extended regularization information criterion. The Z column gives the frequencies for the individual country characteristics; the other columns give the frequencies for the country characteristics interacted with the global variables.

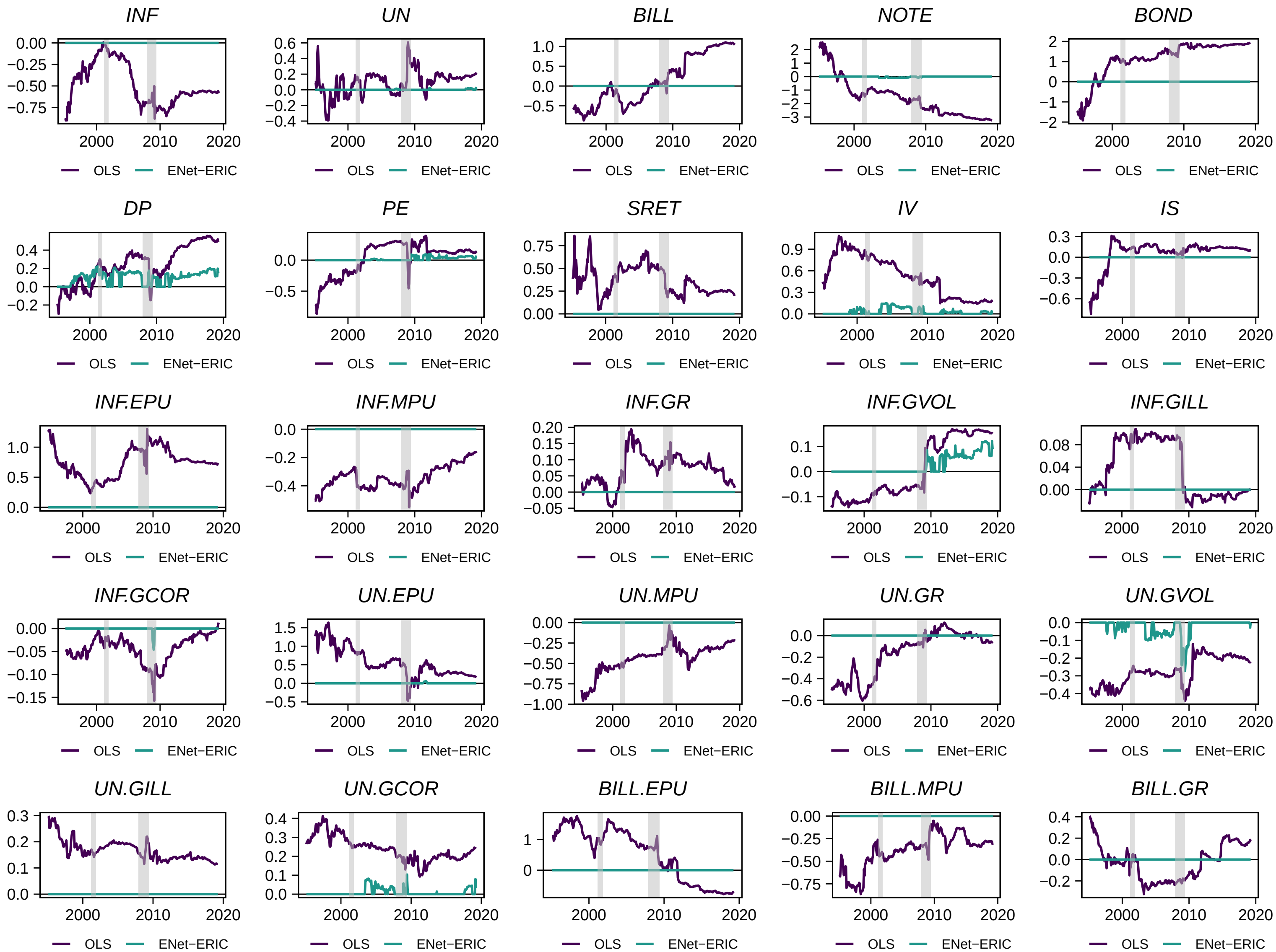


Figure 5: Recursive coefficient estimates

The figure shows recursive ordinary least squares (OLS) and elastic net (ENet) slope coefficient estimates for a panel predictive regression with 70 predictors, where the regularization parameter for the elastic net is tuned via the extended regularization information criterion (ERIC). Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

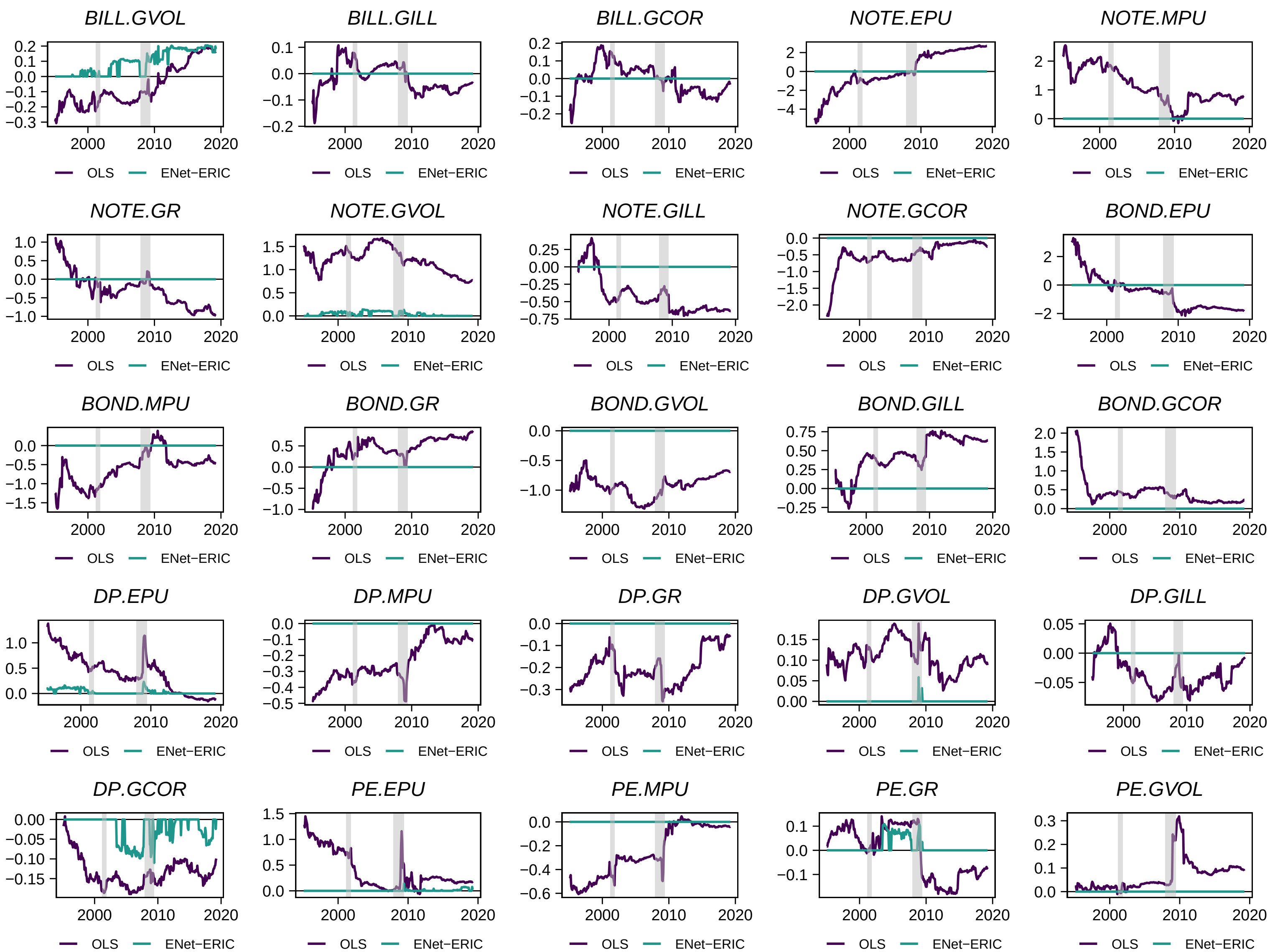


Figure 5 (continued)

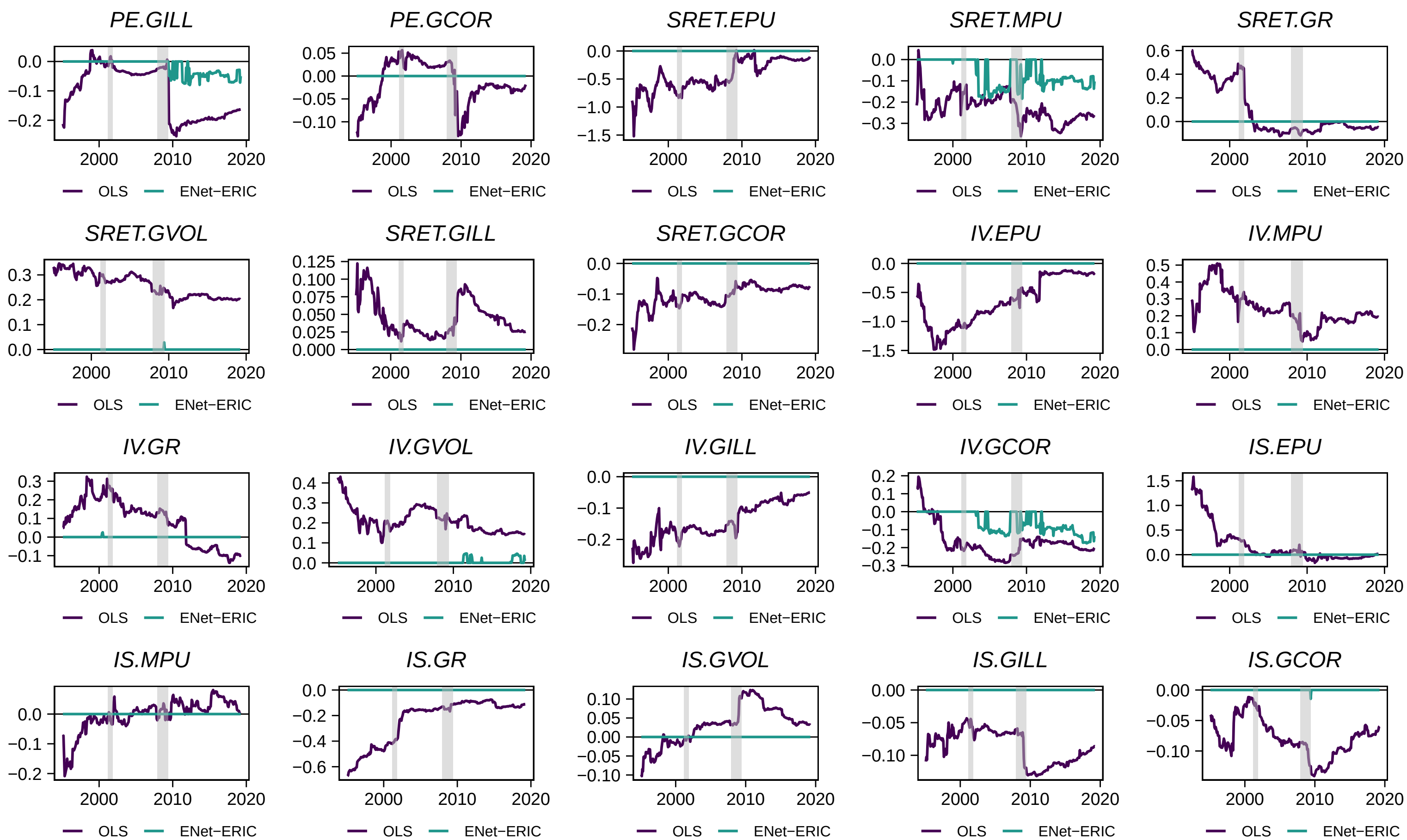


Figure 5 (continued)

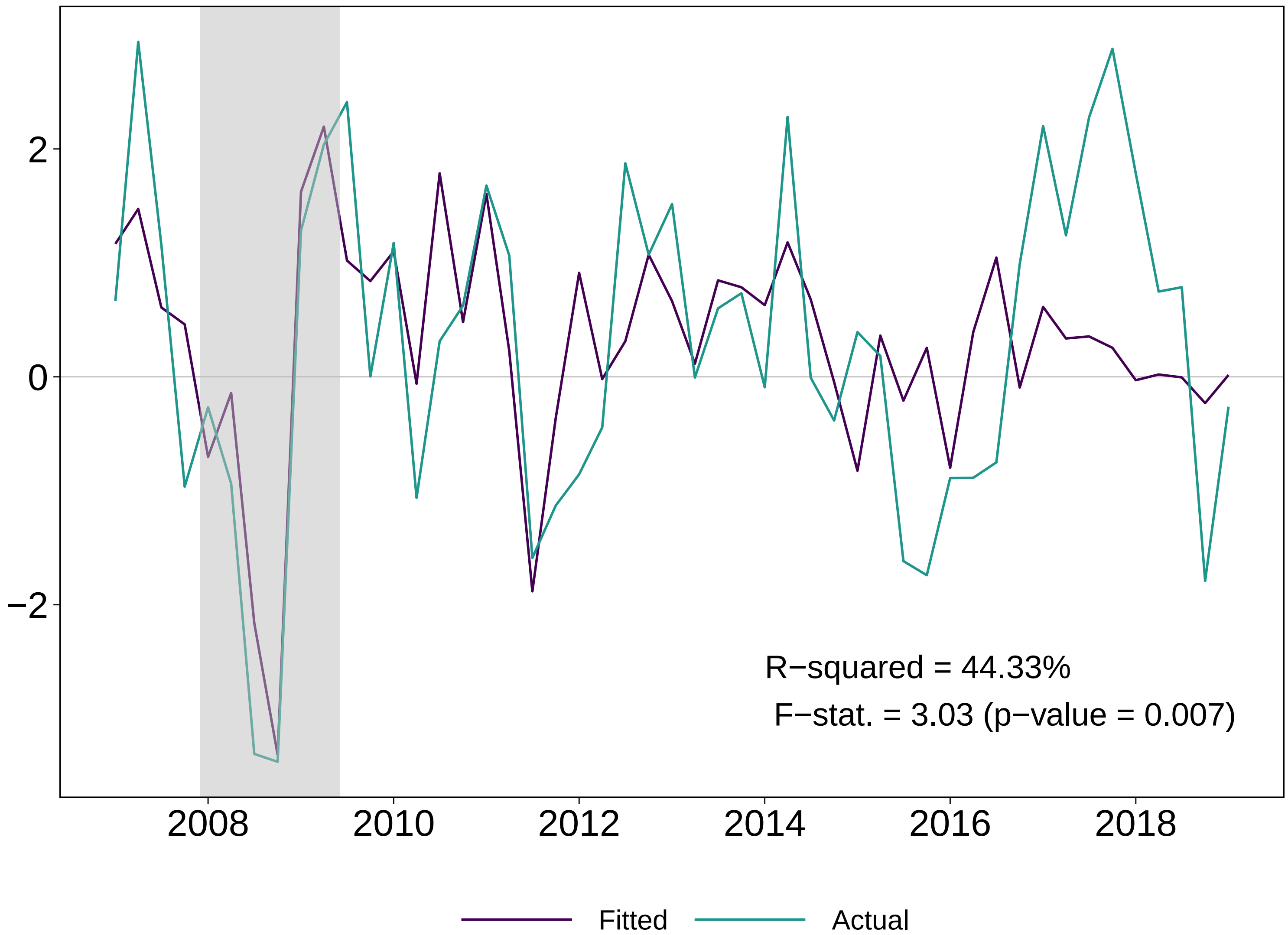


Figure 6: Fitted and Actual Values

The figure shows fitted and actual values for a regression of the Lilley et al. (forthcoming) change in US foreign bond holdings on the ten predictors selected by the elastic net, where the regularization parameter is tuned via the extended regularization information criterion. The change in US foreign bond holdings is quarterly. The monthly predictors are aggregated over the three months comprising a quarter and across countries. The sample is 2007:1 to 2019:2. The variables are standardized before estimating the regression. The vertical bar delineates the most recent business-cycle recession as dated by the National Bureau of Economic Research.

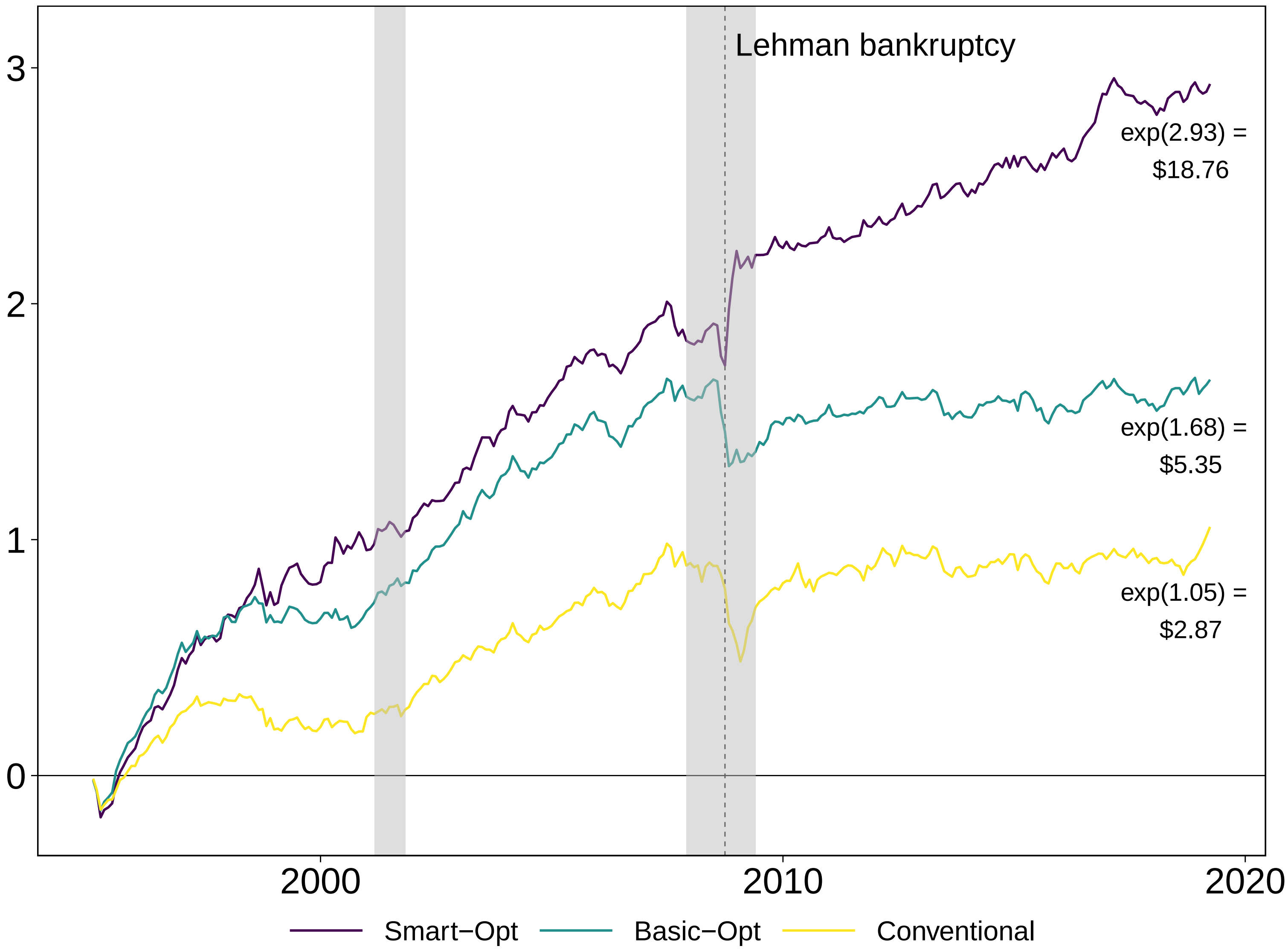


Figure 7: Log cumulative excess returns

The figure shows log cumulative portfolio excess returns for a US investor with mean-variance preferences and a relative risk aversion coefficient of five who invests in 14 foreign currencies. The Basic-Opt (Smart-Opt) portfolio assumes that the investor uses the no-change (elastic net) exchange rate forecast, where the regularization parameter for the elastic net is tuned via the extended regularization information criterion. The conventional portfolio goes long (short) currencies with the highest (lowest) bill yield differentials. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

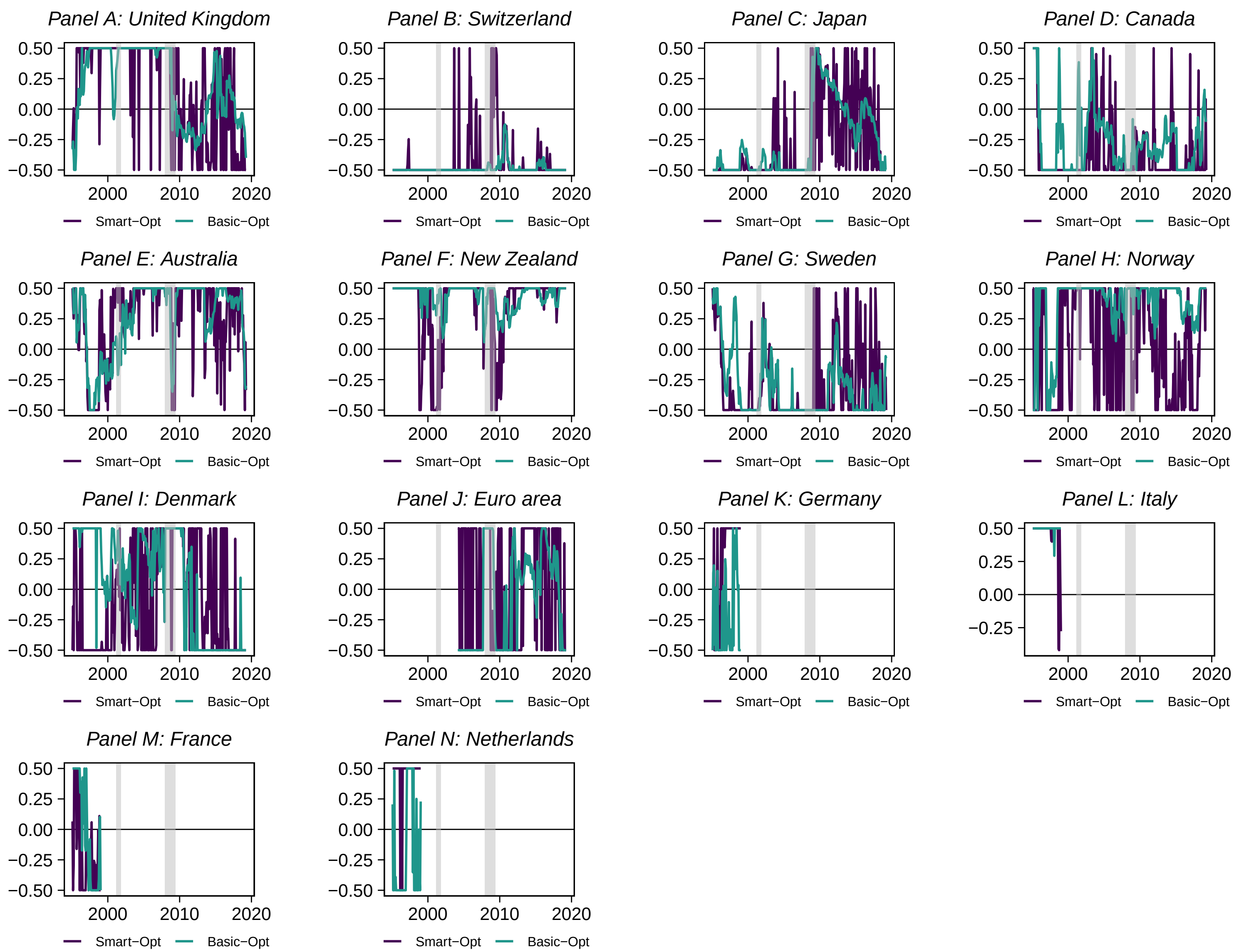


Figure 8: Portfolio weights

The figure shows portfolio weights for a US investor with mean-variance preferences and a relative risk aversion coefficient of five who invests in 14 foreign currencies. The Basic-Opt (Smart-Opt) portfolio assumes that the investor uses the no-change (elastic net) exchange rate forecast, where the regularization parameter for the elastic net is tuned via the extended regularization information criterion. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

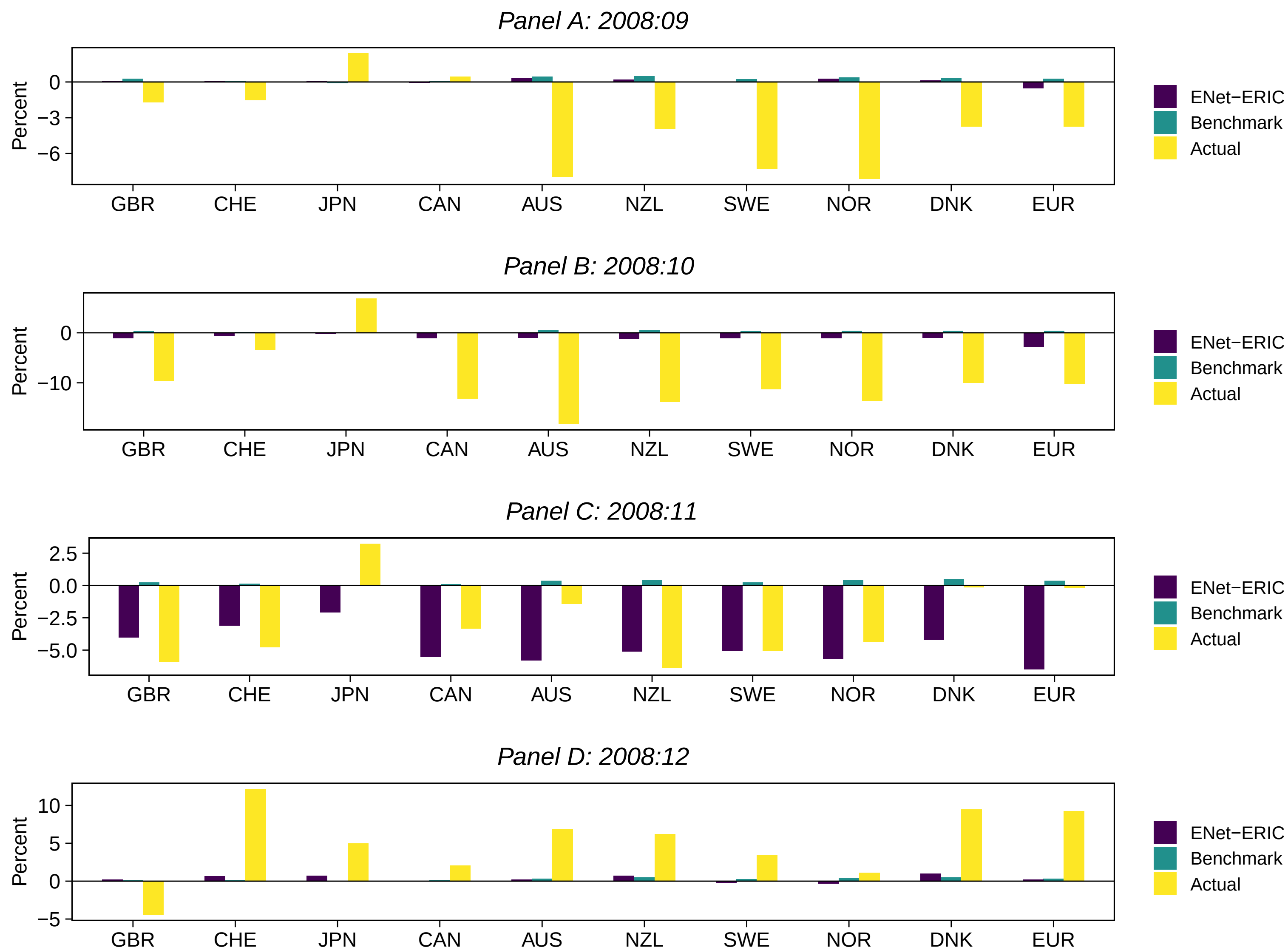


Figure 9: Currency excess return forecasts for late 2008

The figure shows currency excess return forecasts for the final four months of 2008. The ENet (benchmark) forecast uses the elastic net (no-change) exchange rate forecast in conjunction with the bill yield differential to forecast the currency excess return. The regularization parameter for the elastic net is tuned via the extended regularization information criterion (ERIC).

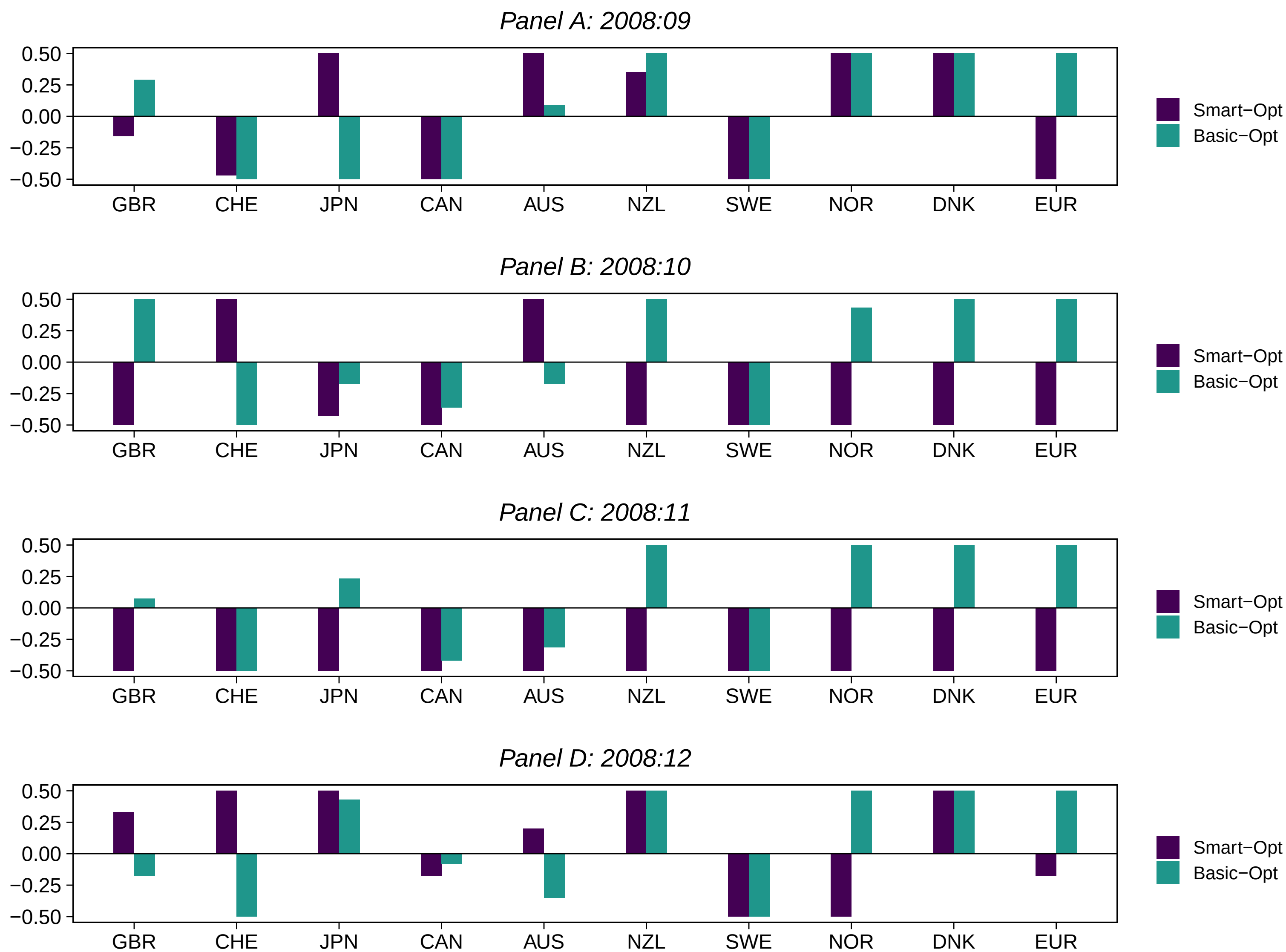


Figure 10: Portfolio weights for late 2008

The figure shows portfolio weights for the final four months of 2008 for a US investor with mean-variance preferences and a relative risk aversion coefficient of five who invests in ten foreign currencies. The Smart-Opt (Basic-Opt) portfolio assumes that the investor uses the elastic net (no-change) exchange rate forecast, where the regularization parameter for the elastic net is tuned via the extended regularization information criterion.

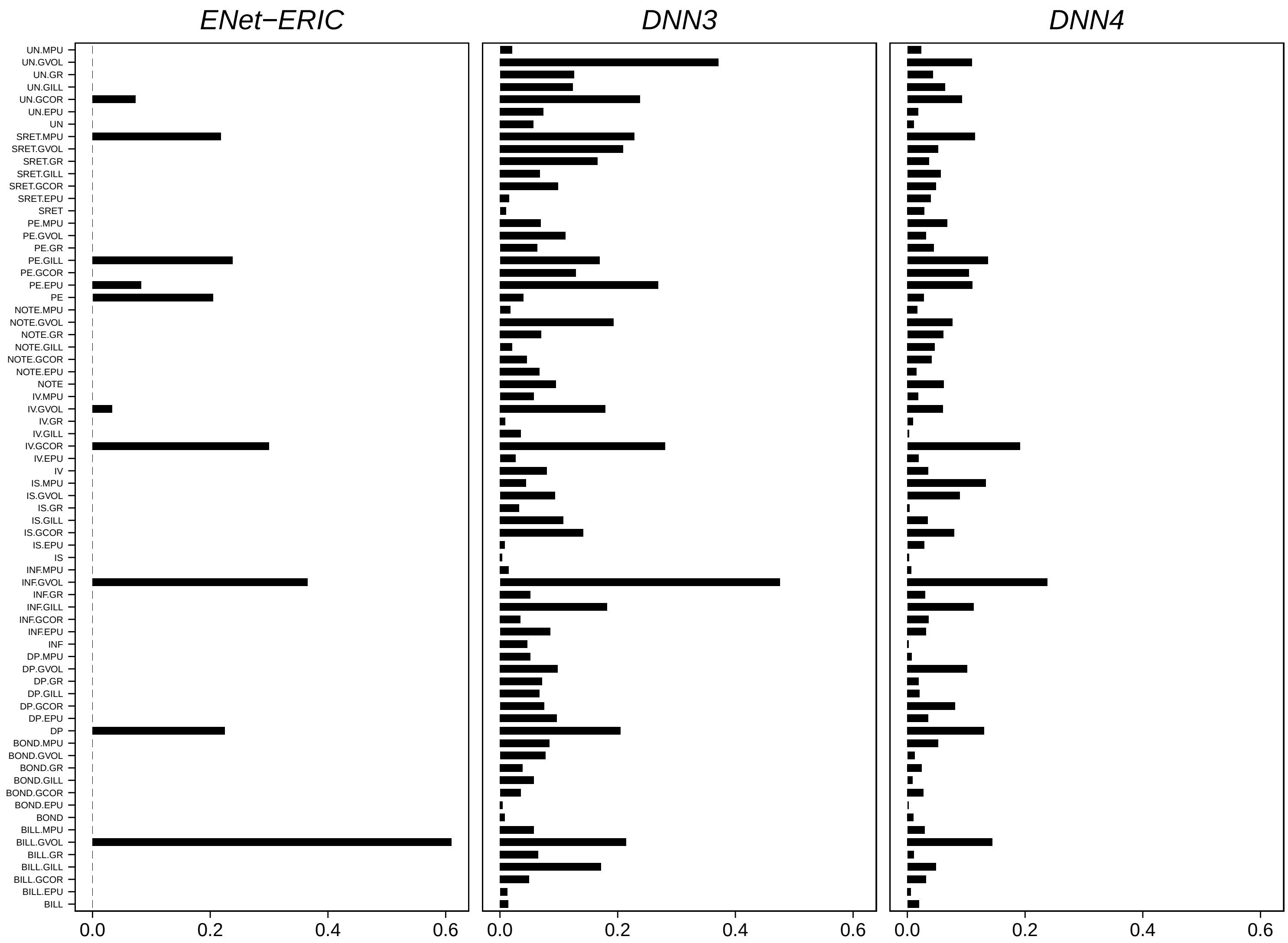


Figure 11: Variable importance

The figure shows variable importance measures for 70 predictors for fitted models used to generate forecasts based on a panel predictive regression estimated via the elastic net (ENet) and deep neural networks with three and four hidden layers (DNN3 and DNN4, resepctively), where the regularization parameter for the elastic net is tuned via the extended regularization information criterion (ERIC).

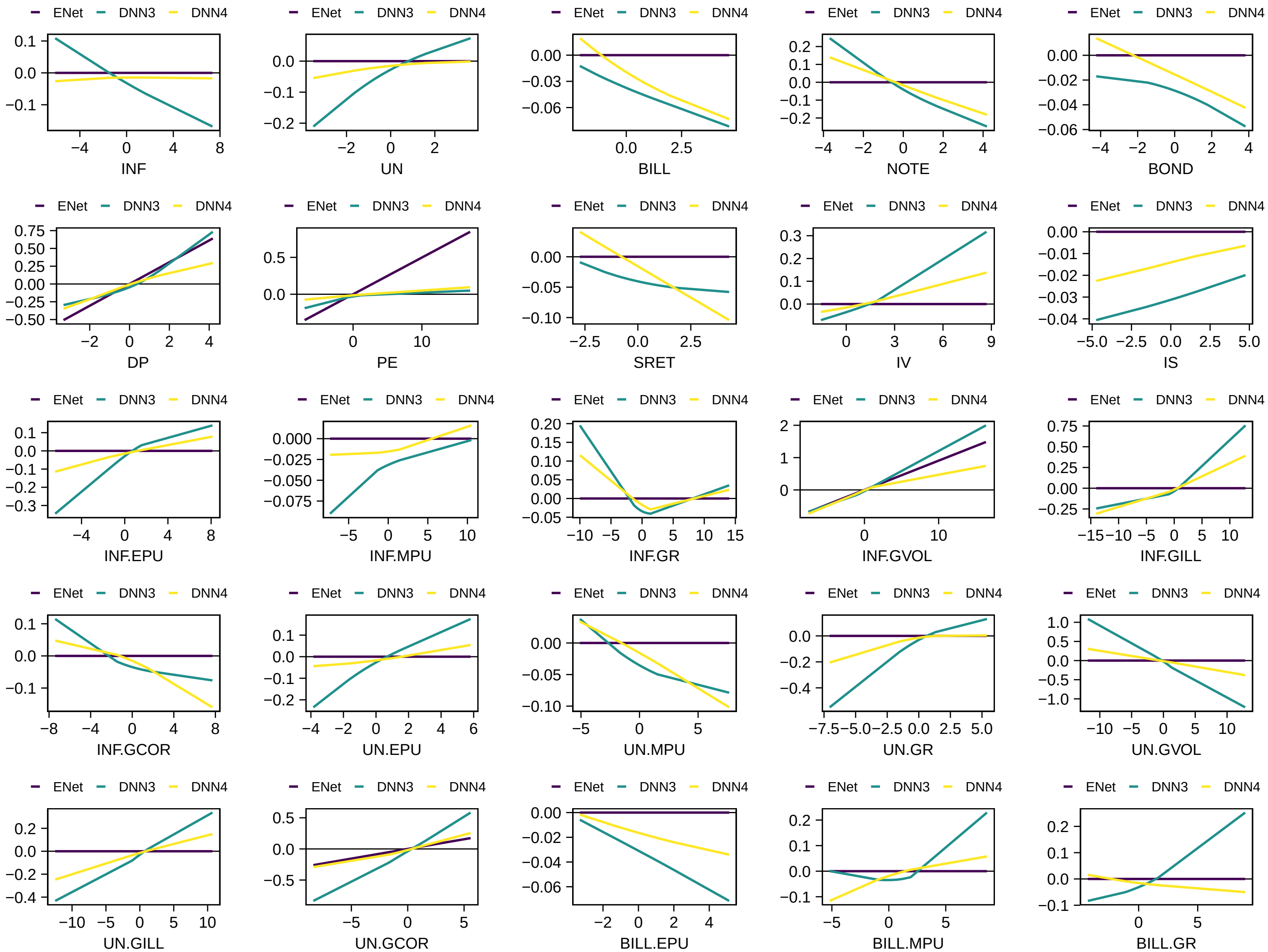


Figure 12: Partial dependence plots

The figure shows partial dependence plots for 70 predictors for fitted models used to generate forecasts based on a panel predictive regression estimated via the elastic net (ENet) and deep neural networks with three and four hidden layers (DNN3 and DNN4, resepctively), where the regularization parameter for the elastic net is tuned via the extended regularization information criterion.

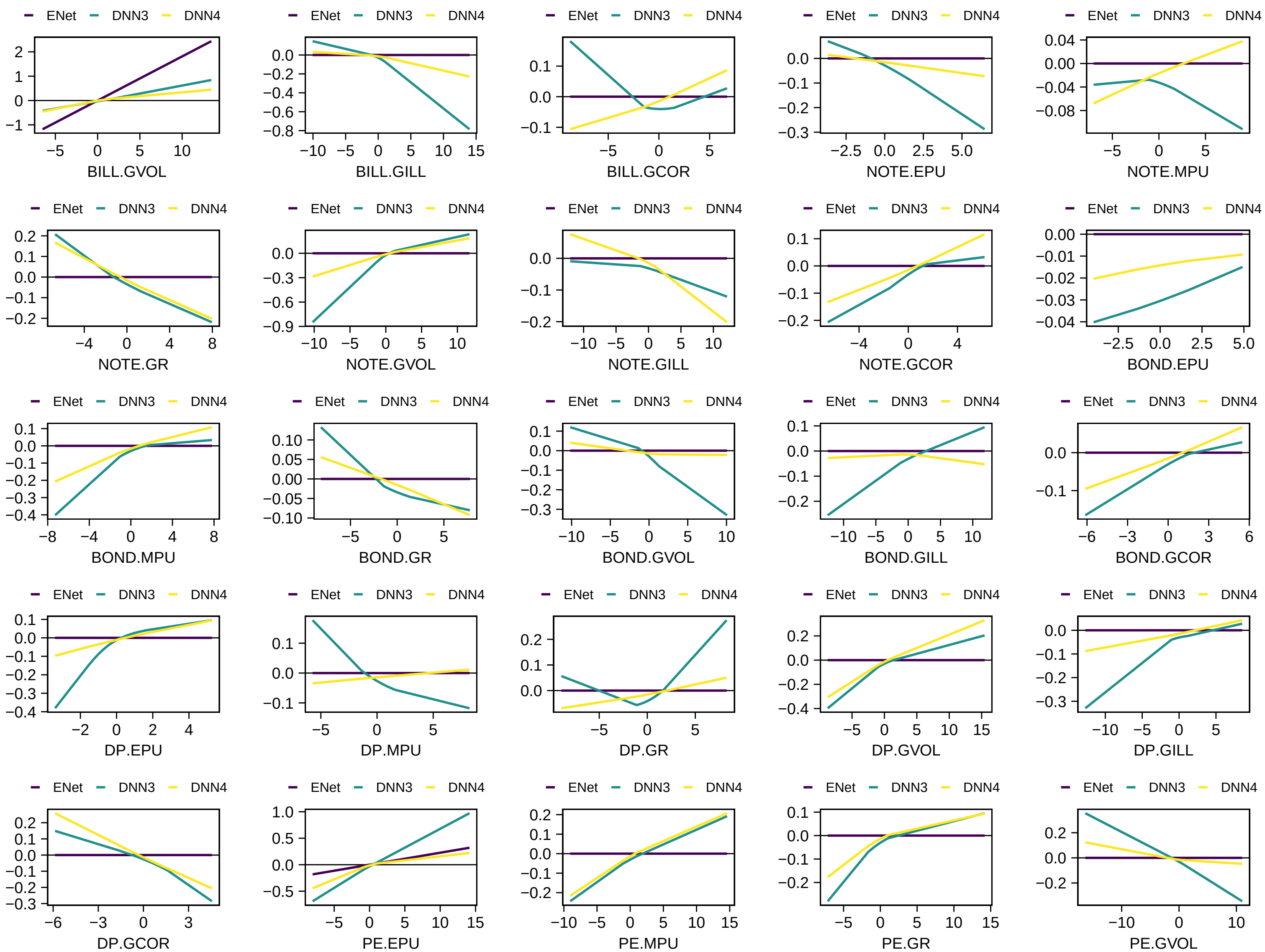


Figure 12 (continued)

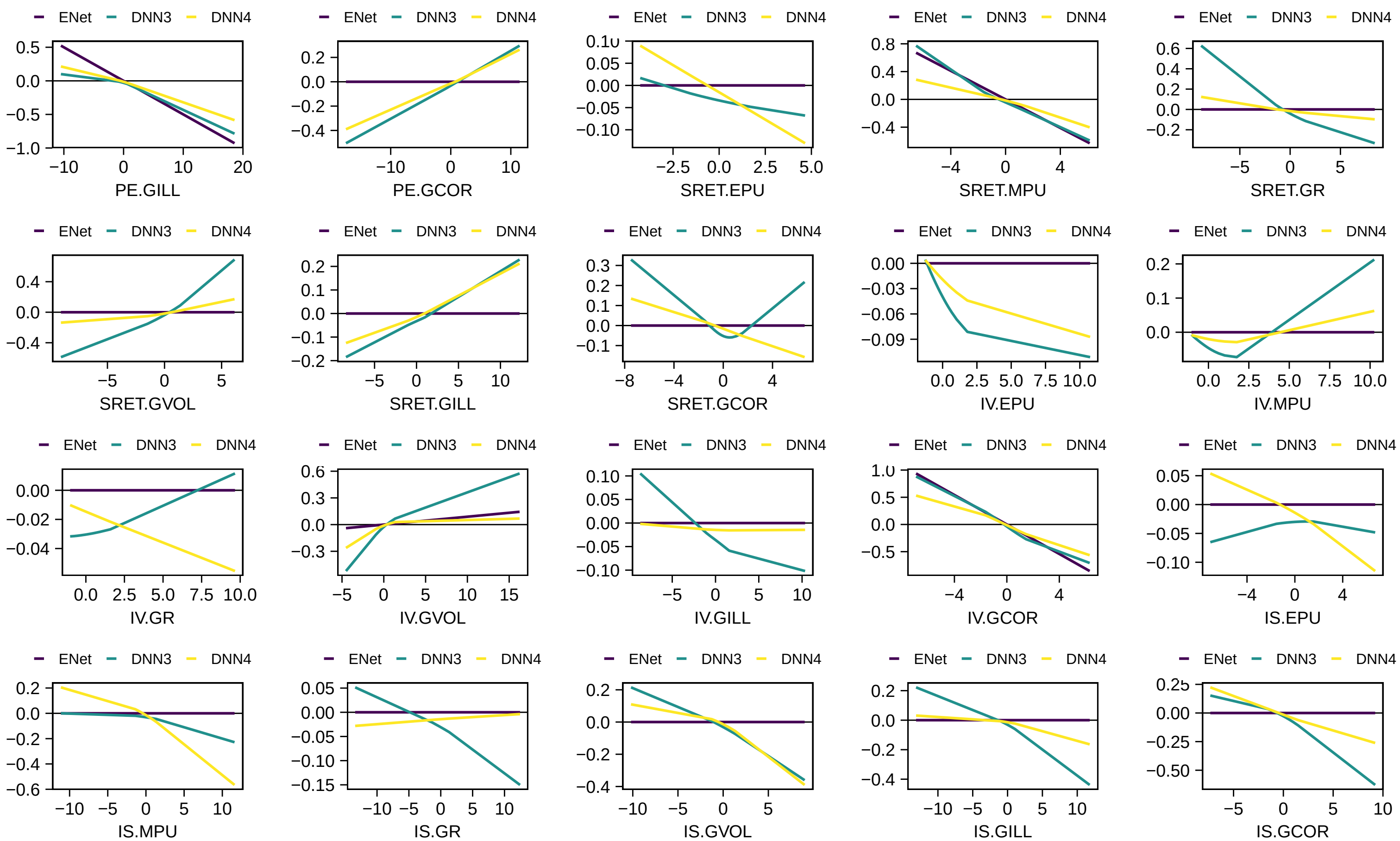


Figure 12 (continued)

Internet Appendix for
"Exchange Rate Prediction with Machine Learning and a Smart Carry Trade Portfolio"

Not for Publication

September 1, 2020

A1 Data

A1.1 Exchange Rates

Daily bid and ask spot and forward exchange rates are from Barclays and Reuters via Datastream. Datastream country mnemonics are as follows: United Kingdom, `GBP`; Switzerland, `CHF`; Japan, `JPY`; Canada, `CAD`; Australia, `AUD`; New Zealand, `NZD`; Sweden, `SEK`; Norway, `NOK`; Denmark, `DKK`; Euro area, `EUR`; Germany, `DEM`; Italy, `ITL`; France, `FRF`; Netherlands, `NLG`.

Bid spot price Ticker `BB***SP(EB)`, where "`***`" indicates the country mnemonic.

Ask spot price Ticker `BB***SP(EO)`.

Bid forward price Ticker `BB***1F(EB)`.

Ask forward price Ticker `BB***1F(EO)`.

A1.2 Country Characteristics

Country characteristics are computed using data from Global Financial Data (GFD) and the Organization for Economic Cooperation and Development (OECD). Country mnemonics are as follows: United Kingdom, GBR; Switzerland, CHE; Japan, JPN; Canada, CAN; Australia, AUS; New Zealand, NZL; Sweden, SWE; Norway, NOR; Denmark, DNK; Euro area, EUR; Germany, DEU; Italy, ITA; France, FRA; Netherlands, NLD; United States, USA.

Inflation differential (INF) Inflation rates are computed from consumer price index (CPI) data from GFD (ticker CP***M); CPI data for the Euro area (EA19) are from the OECD (available at https://data.oecd.org/price/inflation-cpi.htm).

Unemployment gap differential (UN) Unemployment rates are from GFD (ticker UN***M); the unemployment rate for the Euro area (EA19) is from the OECD (available at https://data.oecd.org/unemp/unemployment-rate.htm).

Bill yield differential (BILL) Bill yields are three-month Treasury bill yields from GFD (ticker IT***3D; for the Euro area, IBEUR3D).

Note yield differential (NOTE) Note yields are five-year government bond yields from GFD (ticker IG***5D).

Bond yield differential (BOND) Bond yields are ten-year government bond yield data from GFD (ticker IG***10D).

Dividend yield differential (DP) Dividend yields are from GFD (ticker SY***YM; for the United Kingdom, _DFTASD, for Canada, SYCANYTM; for the Netherlands, SYNLDYAM).

Price-earnings differential (PE) Price-earnings ratios are from GFD (ticker SY***PM; for the United Kingdom, _PFTASD; for Japan, SYJPNPTM; for Canada, SYCANPTM).

Stock market momentum (SRET) Twelve-month cumulative returns are computed from total return indices from GFD (tickers are as follows: United Kingdom, _TFTASD; Switzerland, _SSHID; Japan, _TOPXDVD; Canada, _TRGSPTSE; Australia, _AORDAD; New Zealand, _NZGID; Sweden, _OMXSBGI; Norway, _OSEAXD; Denmark, _OMXCGID; Euro area, `_DMIEUOD`; Germany, `_CDAXD`; Italy, `_BCIPRD`; France, `TRSBF250D`; Netherlands, `_AAXGRD`; United States, `_SPXTRD`).

Idiosyncratic volatility (IV) Following Filippou, Gozluklu, and Taylor (2018), we compute daily currency excess returns using daily spot and forward rates from Datastream for the 14 countries that we analyze. We construct a daily dollar risk factor ($\mathrm{MKT}_{\mathrm{FX}}$) as the cross-sectional average of the daily currency excess returns. To construct a daily carry trade risk factor ($\mathrm{HML}_{\mathrm{FX}}$), we first create six portfolios by sorting on the previous daily forward discount; the carry trade factor is the return for the long-short portfolio. Each month, we regress daily currency excess returns for country i on a constant and the $\mathrm{MKT}_{\mathrm{FX}}$ and $\mathrm{HML}_{\mathrm{FX}}$ factors:

$$RX^i_{t,d} = \alpha^i + \beta^i_{\mathrm{MKT}_{\mathrm{FX}},t}\mathrm{MKT}_{\mathrm{FX},t,d} + \beta^i_{\mathrm{HML}_{\mathrm{FX}},t}\mathrm{HML}_{\mathrm{FX},t,d} + \varepsilon^i_{t,d}, \tag{A1.1}$$

where $RX^i_{t,d}$ is the day-d currency excess return for country i for month t and $\mathrm{MKT}_{\mathrm{FX},t,d}$ ($\mathrm{HML}_{\mathrm{FX},t,d}$) is the day-$d$ return for month t for the dollar (carry trade) factor. Idiosyncratic volatility is defined as

$$\mathrm{IV}_{i,t} = \left[\frac{1}{T_{i,t}}\sum_{d=1}^{T_{i,t}}\left(\hat{\varepsilon}^i_{t,d}\right)^2\right]^{0.5}, \tag{A1.2}$$

where $\hat{\varepsilon}^i_{t,d}$ is the fitted ordinary least squares residual for Equation (A1.1) and $T_{i,t}$ is number of daily currency excess return observations available for country i for month t.

Idiosyncratic skewness (IS) Defined as

$$\mathrm{IS}_{i,t} = \left(\frac{1}{T_{i,t}-2}\right)\frac{\sum_{d=1}^{T_{i,t}}\left(\hat{\varepsilon}^i_{t,d}\right)^3}{\left(\mathrm{IV}_{i,t}\right)^3}, \tag{A1.3}$$

where $\text{IV}_{i,t}$ is given by Equation (A1.2).[1]

A1.3 Global Variables

Economic policy uncertainty (EPU) Available from the Economic Policy Uncertainty website at `https://www.policyuncertainty.com/us_monthly.html`.

Monetary policy uncertainty (MPU) Available from the Economic Policy Uncertainty website at `https://www.policyuncertainty.com/monetary.html`.

Geopolitical risk (GR) Available from the Economic Policy Uncertainty website at `https://www.policyuncertainty.com/gpr.html`.

Global foreign exchange volatility (GVOL) As in Menkhoff et al. (2012), month-t global foregin exchange volatility is defined as

$$\text{GVOL}_t = \frac{1}{T_t} \sum_{d \in T_t} \left[\sum_{k \in K_{t,d}} \left(\frac{|\Delta s^k_{t,d}|}{K_{t,d}} \right) \right], \tag{A1.4}$$

where $\Delta s^k_{t,d}$ is the day-d change in the log exchange rate for country k and month t, T_t is the number of days in month t, and $K_{t,d}$ is the number of currencies available for day d in month t.

Global foreign exchange illiquidity (GILL) As in Menkhoff et al. (2012), month-t global foreign exchange illiquidity is defined as

$$\text{GILL}_t = \frac{1}{T_t} \sum_{d \in T_t} \left[\sum_{k \in K_{t,d}} \left(\frac{|\text{BAS}^k_{t,d}|}{K_{t,d}} \right) \right], \tag{A1.5}$$

where $\text{BAS}^k_{t,d}$ is the day-d bid-ask exchange rate spread (in percent) for country k and month t.

[1] The construction of idiosyncratic volatility and skewness follows Goyal and Santa-Clara (2003), Fu (2009), Boyer, Mitton, and Vorkink (2010), and Chen and Petkova (2012).

Global foreign exchange correlation (GCOR) Similarly to Mueller, Stathopoulos, and Vedolin (2017), month-t foreign exchange currency correlation is defined as

$$\mathrm{GCOR}_t = \frac{1}{N_t^{\mathrm{comb}}} \sum_{i=1}^{N_t} \left(\sum_{j>i} \mathrm{RC}_t^{i,j} \right), \tag{A1.6}$$

where $\mathrm{RC}_t^{i,j}$ is the realized correlation between currency excess returns for countries i and j based on daily data for month t, N_t^{comb} is the number of combinations of currencies (i, j), and N_t is the number of available currencies.

A2 Training the Deep Neural Networks

We estimate the deep neural networks (DNNs) in `Python` using the `keras` package. In implementing the Adam algorithm (Kingma and Ba 2015) to estimate the weights of the DNNs, we use the following hyperparameter values:

- *dropout rate* (Hinton et al. 2012; Srivastava et al. 2014) of 0.5 for each hidden layer;
- shrinkage parameter of 0.01 for ℓ_1 regularization;
- reduce the *learning rate* by a factor of 0.05 when the validation loss stops improving;
- *epochs* equal 100;
- *batch size* of 32;
- training (validation) sample split of 0.8 (0.2).

We also use *batch normalization* (Ioffe and Szegedy 2015). The hyperparameter values are designed to better guard against overfitting.

References

Bao, Y. (2009). Estimation Risk-Adjusted Sharpe Ratio and Fund Performance Ranking under a General Return Distribution. *Journal of Financial Econometrics* 7:2, 152–173.

Boyer, B., T. Mitton, and K. Vorkink (2010). Expected Idiosyncratic Skewness. *Review of Financial Studies* 23:1, 169–202.

Chen, Z. and R. Petkova (2012). Does Idiosyncratic Volatility Proxy for Risk Exposure? *Review of Financial Studies* 25:9, 2745–2787.

Clark, T. E. and K. D. West (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics* 138:1, 291–311.

Filippou, I., A. E. Gozluklu, and M. P. Taylor (2018). Global Political Risk and Currency Momentum. *Journal of Financial and Quantitative Analysis* 53:5, 2227–2259.

Fu, F. (2009). Idiosyncratic Risk and the Cross-Section of Expected Stock Returns. *Journal of Financial Economics* 91:1, 24–27.

Goyal, A. and P. Santa-Clara (2003). Idiosyncratic Risk Matters! *Journal of Finance* 58:3, 975–1007.

Hinton, G. E., N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov (2012). Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors. Working paper (arXiv:1207.0580).

Ioffe, S. and C. Szegedy (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *Proceedings of the 32nd International Conference on Machine Learning.*

Kingma, D. P. and J. Ba (2015). *Adam: A Method for Stochastic Optimization.* Third Annual International Conference on Learning Representations. San Diego.

Lustig, H., N. Roussanov, and A. Verdelhan (2011). Common Risk Factors in Currency Markets. *Review of Financial Studies* 24:11, 3731–3777.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012). Carry Trades and Global Foreign Exchange Volatility. *Journal of Finance* 67:2, 681–718.

Mueller, P., A. Stathopoulos, and A. Vedolin (2017). International Correlation Risk. *Journal of Financial Economics* 126:2, 270–299.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15:56, 1929–1958.

Table A1: R^2_{OS} statistics (in percent), five-fold cross validation and AICc tuning

(1)	(2)	(3)	(4)	(5)
Country	Obs.	OLS	ENet-CV	ENet-AICc
United Kingdom	291	-10.83	-0.24	-5.19
Switzerland	291	-11.78	-2.08	-7.23
Japan	291	-10.53	-2.93	-8.08
Canada	291	-18.76	-8.99	-13.02
Australia	291	-3.57	-2.17	-2.68
New Zealand	291	-5.41	-0.65	-4.38
Sweden	291	-2.60	-1.05	-1.07
Norway	291	-1.88	-1.00	-1.23
Denmark	291	-6.22	0.55	-3.31
Euro area	230	-10.76	-4.18	-10.38
Germany	48	-11.59	-5.78	-4.50
Italy	48	-12.53	-5.09	-9.06
France	48	-36.64	-16.34	-18.31
Netherlands	48	-21.81	-7.04	-12.78
All	3,041	-8.03	-2.32	-5.40

The table reports out-of-sample R^2 (R^2_{OS}) statistics for monthly out-of-sample forecasts of exchange rate changes based on a country-level predictive regression estimated via ordinary least squares (OLS) and the elastic net (ENet), where the regularization parameter for the elastic net is tuned via five-fold cross validation (CV) or the corrected AIC (AICc). The R^2_{OS} statistic is the percent reduction in mean squared forecast error (MSFE) for a competing forecast vis-á-vis the no-change benchmark forecast. The competing forecast is based on a country-level predictive regression with 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables. For a positive R^2_{OS} statistic, we use the Clark and West (2007) MSFE-adj statistic to test the null hypothesis that the benchmark MSFE is less than or equal to the competing MSFE against the alternative hypothesis that the benchmark MSFE is greater than the competing MSFE; 0.00 indicates less than 0.005 in absolute value; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The second column reports the number of out-of-sample observations.

Table A2: R^2_{OS} statistics (in percent), country-level estimation and all predictors

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Country	Obs.	OLS	ENet-CV	ENet-AICc	ENet-BIC	ENet-MBIC	ENet-GIC	ENet-ERIC
United Kingdom	291	−211.07	−0.98	0.79	−0.10	–	–	–
Switzerland	291	−184.96	0.91**	−4.23	–	–	–	–
Japan	291	−122.88	−0.30	−0.77	–	–	–	–
Canada	291	−67.50	−0.64	−1.66	0.35	0.39	0.43	0.41
Australia	291	−121.89	1.92***	−1.70	0.60**	0.26**	0.67***	0.07*
New Zealand	291	−521.56	−3.87	−1.60	–	–	–	–
Sweden	291	−86.74	0.14	−2.43	–	–	–	–
Norway	291	−92.52	0.03	2.43**	−0.14	−0.02	−0.08	−0.02
Denmark	291	−98.33	−4.18	−6.84	−0.11	–	−0.11	–
Euro area	122	−411.34	−1.68	−5.58	−4.09	–	–	–
Germany	48	−168.39	0.00	−1.08	–	–	–	–
Italy	48	−724.13	−0.86	−0.81	–	–	–	–
France	48	−356.45	0.88	−5.14	–	–	–	–
Netherlands	48	−521.56	−3.87	−1.60	–	–	–	–

The table reports out-of-sample R^2 (R^2_{OS}) statistics for monthly out-of-sample forecasts of exchange rate changes based on a country-level predictive regression estimated via ordinary least squares (OLS) and the elastic net (ENet). The regularization parameter for the elastic net is tuned via five-fold cross validation (CV), the corrected AIC (AICc), Bayesian information criterion (BIC), modified BIC (MBIC), generalized information criterion (GIC), or extended regularization information criterion (ERIC). The R^2_{OS} statistic is the percent reduction in mean squared forecast error (MSFE) for a competing forecast vis-á-vis the no-change benchmark forecast. The competing forecast is based on a country-level predictive regression with 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables. For a positive R^2_{OS} statistic, we use the Clark and West (2007) MSFE-adj statistic to test the null hypothesis that the benchmark MSFE is less than or equal to the competing MSFE against the alternative hypothesis that the benchmark MSFE is greater than the competing MSFE; 0.00 indicates less than 0.005 in absolute value; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively; – indicates that no predictor is ever selected by the elastic net. The second column reports the number of out-of-sample observations.

Table A3: R^2_{OS} statistics (in percent), panel estimation and individual characteristics

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Country	Obs.	INF	UN	BILL	NOTE	BOND	DP	PE	SRET	IV	IS	CPI	TR
United Kingdom	291	−0.19	−0.42	−0.34	−0.03	0.02	0.14	−0.42	1.12**	0.94**	−0.04	−1.44	−0.64
Switzerland	291	−0.04	−0.02	−0.05	0.39	0.22	−3.05	0.09	0.33	0.93	−0.13	−1.29	−0.05
Japan	291	−0.01	0.42*	−0.32	−0.29	−0.29	−2.09	0.06	0.04	−0.62	0.01	−0.25	0.42*
Canada	291	−0.06	0.54	−0.42	−0.12	−0.06	−2.77	0.06	−0.38	−2.23	−0.02	−0.43	0.46
Australia	291	−0.40	0.59*	0.04	0.36*	0.12	−0.72	0.07	0.92**	−0.46	0.00	0.04	0.19
New Zealand	291	−0.71	0.44*	−0.03	0.08	0.16	0.21	−0.25	1.02***	0.08	−0.03	0.38	−0.28
Sweden	291	−0.09	0.15	−0.13	0.17	−0.04	−1.03	0.06	0.16	−0.80	−0.08	0.17	0.04
Norway	291	−0.25	0.18	−0.09	0.24	0.13	−0.88	0.05	−0.05	−1.19	−0.11	0.31	−0.02
Denmark	291	−0.29	−0.27	−0.38	−0.07	−0.15	−1.48	0.03	0.08	1.19**	−0.03	0.00	−0.53
Euro area	230	−0.10	−0.22	−0.20	−0.40	−0.08	−1.65	−0.10	0.44	0.20	−0.05	0.01	−0.24
Germany	48	0.39	−0.37	−0.80	−0.14	−0.01	−0.30	−1.77	0.41	3.06*	−0.23	0.54	−0.17
Italy	48	−0.49	−0.18	−2.74	−2.10	−1.96	−0.03	0.18	−1.23	−1.48	0.63	−2.21	−0.41
France	48	−0.79	−0.72	−1.26	−0.94	−1.26	−4.09	0.96	−0.80	1.59	−1.16	−0.44	−1.38
Netherlands	48	−0.69	−0.26	−1.11	−0.23	−0.12	−0.23	0.84	−3.73	3.55**	−0.24	0.33	−0.88
All	3,041	−0.26	0.16**	−0.22	0.03*	−0.02	−1.23	−0.02	0.33***	−0.10	−0.06	−0.17	−0.08

The table reports out-of-sample R^2 (R^2_{OS}) statistics for monthly out-of-sample forecasts of exchange rate changes based on a panel predictive regression estimated via ordinary least squares. Each predictive regression uses the characteristic in the column heading as a predictor; CPI is the difference between the country-i and US log consumer price index; TR indicates that the predictive regression uses INF and UN. The R^2_{OS} statistic is the percent reduction in mean squared forecast error (MSFE) for a competing forecast vis-á-vis the no-change benchmark forecast. For a positive R^2_{OS} statistic, we use the Clark and West (2007) MSFE-adj statistic to test the null hypothesis that the benchmark MSFE is less than or equal to the competing MSFE against the alternative hypothesis that the benchmark MSFE is greater than the competing MSFE; 0.00 indicates less than 0.005 in absolute value; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The second column reports the number of out-of-sample observations.

Table A4: R^2_{OS} statistics (in percent), country-level estimation and individual characteristics

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Country	Obs.	INF	UN	BILL	NOTE	BOND	DP	PE	SRET	IV	IS	CPI	TR
United Kingdom	291	−0.42	−0.94	−1.30	−0.29	−0.24	−0.20	−2.37	0.95**	0.50*	−0.61	−0.86	−1.52
Switzerland	291	−1.66	−0.51	0.05	−0.24	−0.54	−4.86	−1.42	−0.01	0.12	−0.28	−5.25	−2.28
Japan	291	−0.24	0.13	−0.79	−1.30	−0.85	−2.67	−2.30	−0.40	−2.00	−0.64	−4.76	−0.04
Canada	291	0.27	0.28	−0.82	−0.67	−0.53	−0.71	−0.37	−0.43	−0.36	0.22	−0.88	0.62
Australia	291	−0.90	0.49	0.19	0.68*	0.41	−0.54	0.33	0.60**	−0.29	−0.10	−0.37	−0.46
New Zealand	291	−2.08	0.32	−0.25	−0.21	−0.11	−0.08	−0.35	1.60***	−0.92	−0.28	−0.57	−1.78
Sweden	291	−1.63	0.05	−2.15	−0.16	−0.14	−4.04	−0.29	−0.02	−1.35	−0.21	−2.18	−1.58
Norway	291	0.04	−0.04	−0.96	0.25	0.28	−0.12	0.01	−0.18	−0.63	−1.15	−0.88	0.01
Denmark	291	−1.28	−0.35	−0.49	−0.41	−1.13	−1.13	−0.10	−0.38	−0.01	−0.86	−1.94	−1.59
Euro area	122	0.77	−0.22	−0.14	−2.52	−2.71	−0.99	0.84	−1.03	−0.69	−1.58	−0.42	1.06*
Germany	48	1.56	0.33	−1.56	−1.01	−0.24	−1.52	2.16	−0.31	3.98*	−0.45	−0.50	1.88
Italy	48	−0.64	−1.13	−24.29	−14.71	−10.27	−2.59	−0.05	−6.64	−3.69	0.03	−9.02	−1.89
France	48	−0.90	−0.32	−1.58	−0.19	−0.70	−6.14	−1.48	−1.36	0.88	6.07*	−6.88	−1.10
Netherlands	48	−0.38	−1.74	−0.90	−0.67	−0.40	−0.06	−0.37	−1.42	5.20**	−2.61	−1.17	−1.93

The table reports out-of-sample R^2 (R^2_{OS}) statistics for monthly out-of-sample forecasts of exchange rate changes based on a country-level predictive regression estimated via ordinary least squares. Each predictive regression uses the characteristic in the column heading as a predictor; CPI is the difference between the country-i and US log consumer price index; TR indicates that the predictive regression uses INF and UN. The R^2_{OS} statistic is the percent reduction in mean squared forecast error (MSFE) for a competing forecast vis-á-vis the no-change benchmark forecast. For a positive R^2_{OS} statistic, we use the Clark and West (2007) MSFE-adj statistic to test the null hypothesis that the benchmark MSFE is less than or equal to the competing MSFE against the alternative hypothesis that the benchmark MSFE is greater than the competing MSFE; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The second column reports the number of out-of-sample observations.

Table A5: Portfolio performance, ENet-BIC forecast

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Basic-Opt portfolio			Smart-Opt portfolio		
Transaction costs	Annualized mean	Annualized volatility	Annualized Sharpe ratio	Annualized mean	Annualized volatility	Annualized Sharpe ratio
Panel A: Full sample (1995:01 to 2019:03)						
Mid-quotes	7.47%	10.22%	0.73***	13.01%	13.40%	0.97***
25%	5.69%	10.13%	0.56***	10.48%	13.40%	0.78***
50%	4.92%	10.11%	0.49**	9.62%	13.39%	0.72***
75%	4.16%	10.11%	0.41**	8.76%	13.39%	0.65***
Bid-ask spread	3.39%	10.10%	0.34*	7.91%	13.38%	0.59***
Panel B: Pre-crisis subsample (1995:01 to 2008:08)						
Mid-quotes	11.88%	10.56%	1.13***	13.81%	12.35%	1.12***
25%	9.80%	10.58%	0.93***	10.79%	12.62%	0.85***
50%	8.81%	10.59%	0.83***	9.70%	12.62%	0.77***
75%	7.82%	10.59%	0.74***	8.62%	12.61%	0.68**
Bid-ask spread	6.83%	10.59%	0.64**	7.53%	12.61%	0.60**
Panel C: Post-crisis subsample (2008:09 to 2019:03)						
Mid-quotes	1.76%	9.55%	0.18	11.99%	14.69%	0.82***
25%	0.39%	9.32%	0.04	10.09%	14.39%	0.70**
50%	−0.09%	9.32%	−0.01	9.52%	14.38%	0.66**
75%	−0.57%	9.31%	−0.06	8.95%	14.38%	0.62**
Bid-ask spread	−1.04%	9.31%	−0.11	8.39%	14.37%	0.58*

The table reports annualized summary statistics for the portfolio excess return for a US investor with mean-variance preferences and a relative risk aversion coefficient of five who invests in 14 foreign currencies. The Basic-Opt (Smart-Opt) portfolio assumes that the investor uses the no-change (elastic net, ENet) exchange rate forecast. The ENet forecast is based on ENet estimation of a panel predictive regression with 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables, where the regularization parameter is tuned via the Bayesian information criterion (BIC). Mid-quotes ignore transaction costs. We account for transaction costs using bid-ask spreads from Datastream. Because the full bid-ask spreads likely overstate transaction costs, we compute results assuming 25%, 50%, and 75% of the full bid-ask spreads. We test the significance of the Sharpe ratios using t-statistics based on Bao (2009) standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table A6: Portfolio performance, ENet-MBIC forecast

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Basic-Opt portfolio			Smart-Opt portfolio		
Transaction costs	Annualized mean	Annualized volatility	Annualized Sharpe ratio	Annualized mean	Annualized volatility	Annualized Sharpe ratio
Panel A: Full sample (1995:01 to 2019:03)						
Mid-quotes	7.47%	10.22%	0.73***	11.70%	13.04%	0.90***
25%	5.69%	10.13%	0.56***	9.40%	13.06%	0.72***
50%	4.92%	10.11%	0.49**	8.57%	13.06%	0.66***
75%	4.16%	10.11%	0.41**	7.75%	13.05%	0.59***
Bid-ask spread	3.39%	10.10%	0.34*	6.92%	13.05%	0.53***
Panel B: Pre-crisis subsample (1995:01 to 2008:08)						
Mid-quotes	11.88%	10.56%	1.13***	13.27%	12.13%	1.09***
25%	9.80%	10.58%	0.93***	10.51%	12.43%	0.85***
50%	8.81%	10.59%	0.83***	9.46%	12.43%	0.76***
75%	7.82%	10.59%	0.74***	8.41%	12.43%	0.68**
Bid-ask spread	6.83%	10.59%	0.64**	7.36%	12.43%	0.59**
Panel C: Post-crisis subsample (2008:09 to 2019:03)						
Mid-quotes	1.76%	9.55%	0.18	9.67%	14.15%	0.68**
25%	0.39%	9.32%	0.04	7.96%	13.87%	0.57*
50%	−0.09%	9.32%	−0.01	7.42%	13.86%	0.54*
75%	−0.57%	9.31%	−0.06	6.89%	13.86%	0.50
Bid-ask spread	−1.04%	9.31%	−0.11	6.35%	13.85%	0.46

The table reports annualized summary statistics for the portfolio excess return for a US investor with mean-variance preferences and a relative risk aversion coefficient of five who invests in 14 foreign currencies. The Basic-Opt (Smart-Opt) portfolio assumes that the investor uses the no-change (elastic net, ENet) exchange rate forecast. The ENet forecast is based on ENet estimation of a panel predictive regression with 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables, where the regularization parameter is tuned via the modified Bayesian information criterion (MBIC). Mid-quotes ignore transaction costs. We account for transaction costs using bid-ask spreads from Datastream. Because the full bid-ask spreads likely overstate transaction costs, we compute results assuming 25%, 50%, and 75% of the full bid-ask spreads. We test the significance of the Sharpe ratios using t-statistics based on Bao (2009) standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table A7: Portfolio performance, ENet-GIC forecast

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Basic-Opt portfolio			Smart-Opt portfolio		
Transaction costs	Annualized mean	Annualized volatility	Annualized Sharpe ratio	Annualized mean	Annualized volatility	Annualized Sharpe ratio
Panel A: Full sample (1995:01 to 2019:03)						
Mid-quotes	7.47%	10.22%	0.73***	12.77%	13.42%	0.95***
25%	5.69%	10.13%	0.56***	10.27%	13.42%	0.77***
50%	4.92%	10.11%	0.49**	9.42%	13.41%	0.70***
75%	4.16%	10.11%	0.41**	8.58%	13.41%	0.64***
Bid-ask spread	3.39%	10.10%	0.34*	7.73%	13.40%	0.58***
Panel B: Pre-crisis subsample (1995:01 to 2008:08)						
Mid-quotes	11.88%	10.56%	1.13***	13.83%	12.46%	1.11***
25%	9.80%	10.58%	0.93***	10.87%	12.74%	0.85***
50%	8.81%	10.59%	0.83***	9.80%	12.74%	0.77***
75%	7.82%	10.59%	0.74***	8.72%	12.73%	0.69**
Bid-ask spread	6.83%	10.59%	0.64**	7.65%	12.73%	0.60**
Panel C: Post-crisis subsample (2008:09 to 2019:03)						
Mid-quotes	1.76%	9.55%	0.18	11.40%	14.60%	0.78**
25%	0.39%	9.32%	0.04	9.50%	14.29%	0.67**
50%	−0.09%	9.32%	−0.01	8.94%	14.29%	0.63**
75%	−0.57%	9.31%	−0.06	8.38%	14.28%	0.59*
Bid-ask spread	−1.04%	9.31%	−0.11	7.82%	14.28%	0.55*

The table reports annualized summary statistics for the portfolio excess return for a US investor with mean-variance preferences and a relative risk aversion coefficient of five who invests in 14 foreign currencies. The Basic-Opt (Smart-Opt) portfolio assumes that the investor uses the no-change (elastic net, ENet) exchange rate forecast. The ENet forecast is based on ENet estimation of a panel predictive regression with 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables, where the regularization parameter is tuned via the generalized information criterion (GIC). Mid-quotes ignore transaction costs. We account for transaction costs using bid-ask spreads from Datastream. Because the full bid-ask spreads likely overstate transaction costs, we compute results assuming 25%, 50%, and 75% of the full bid-ask spreads. We test the significance of the Sharpe ratios using t-statistics based on Bao (2009) standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table A8: Portfolio performance, conventional carry

(1)	(2)	(3)	(4)
Transaction cost	Ann. mean	Ann. volatility	Ann. Sharpe ratio
Panel A: Full sample (1995:01 to 2019:03)			
Mid-quotes	4.84%	9.82%	0.49**
25%	3.98%	9.82%	0.41**
50%	3.58%	9.82%	0.36*
75%	3.18%	9.82%	0.32
Bid-ask spread	2.78%	9.81%	0.28
Panel B: Pre-crisis subsample (1995:01 to 2008:08)			
Mid-quotes	6.65%	8.87%	0.75***
25%	5.70%	8.86%	0.64**
50%	5.22%	8.86%	0.59**
75%	4.74%	8.86%	0.54**
Bid-ask spread	4.27%	8.86%	0.48*
Panel C: Post-crisis subsample (2008:09 to 2019:03)			
Mid-quotes	2.51%	10.94%	0.23
25%	1.76%	10.94%	0.16
50%	1.46%	10.94%	0.13
75%	1.16%	10.94%	0.11
Bid-ask spread	0.86%	10.93%	0.08

The table reports annualized summary statistics for the excess return for a conventional carry trade portfolio. The investor sorts foreign currencies according to the bill yield differential and goes long (short) the fifth (first) quintile portfolio. Mid-quotes ignore transaction costs. We account for transaction costs using bid-ask spreads from Datastream. Because the full bid-ask spreads likely overstate transaction costs, we compute results assuming 25%, 50%, and 75% of the full bid-ask spreads. We test the significance of the Sharpe ratios using t-statistics based on Bao (2009) standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table A9: Portfolio performance, non-US domestic investors

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Basic-Opt portfolio			Smart-Opt portfolio		
Domestic country	Annualized mean	Annualized volatility	Annualized Sharpe ratio	Annualized mean	Annualized volatility	Annualized Sharpe ratio
Panel A: Full sample (1995:01 to 2019:03)						
United Kingdom	6.87%	10.66%	0.64***	10.28%	12.18%	0.84***
Switzerland	7.96%	11.75%	0.68***	8.52%	15.06%	0.57***
Japan	8.79%	11.62%	0.76***	11.24%	14.15%	0.79***
Canada	6.68%	10.55%	0.63***	9.73%	11.93%	0.82***
Australia	7.28%	10.79%	0.67***	9.21%	11.97%	0.77***
New Zealand	5.94%	11.82%	0.50**	11.08%	15.38%	0.72***
Sweden	7.81%	10.17%	0.77***	10.62%	11.70%	0.91***
Norway	7.51%	11.12%	0.68***	11.21%	12.62%	0.89***
Denmark	7.33%	11.18%	0.66***	7.92%	12.12%	0.65***
Panel B: Pre-crisis subsample (1995:01 to 2008:08)						
United Kingdom	11.16%	11.46%	0.97***	12.97%	13.29%	0.98***
Switzerland	13.62%	12.70%	1.07***	11.00%	13.82%	0.80***
Japan	14.54%	13.03%	1.12***	13.70%	14.91%	0.92***
Canada	11.03%	10.97%	1.01***	10.68%	12.01%	0.89***
Australia	11.27%	10.86%	1.04***	11.48%	12.12%	0.95***
New Zealand	9.63%	12.70%	0.76***	10.20%	15.87%	0.64**
Sweden	12.11%	10.30%	1.18***	12.47%	12.26%	1.02***
Norway	13.20%	11.74%	1.12***	14.69%	12.60%	1.17***
Denmark	11.46%	10.15%	1.13***	10.24%	11.34%	0.90***
Panel C: Post-crisis subsample (2008:09 to 2019:03)						
United Kingdom	1.32%	9.33%	0.14	6.80%	10.55%	0.64**
Switzerland	0.65%	10.06%	0.06	5.32%	16.53%	0.32
Japan	1.37%	9.12%	0.15	8.06%	13.09%	0.62*
Canada	1.07%	9.79%	0.11	8.51%	11.86%	0.72**
Australia	2.14%	10.56%	0.20	6.29%	11.76%	0.53*
New Zealand	1.17%	10.48%	0.11	12.20%	14.78%	0.83***
Sweden	2.26%	9.82%	0.23	8.23%	10.94%	0.75**
Norway	0.17%	9.92%	0.02	6.71%	12.58%	0.53*
Denmark	1.99%	12.26%	0.16	4.93%	13.04%	0.38

The table reports annualized summary statistics for the portfolio excess return for a non-US domestic investor with mean-variance preferences and a relative risk aversion coefficient of five who invests in 14 foreign currencies. The Basic-Opt (Smart-Opt) portfolio assumes that the investor uses the no-change (elastic net) exchange rate forecast. The elastic net forecast is based on elastic net estimation of a panel predictive regression with 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with seven global variables, where the regularization parameter is tuned via the extended regularization information criterion. We test the significance of the Sharpe ratios using t-statistics based on Bao (2009) standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table A10: Alphas and currency factor exposures, non-US domestic investors

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Basic-Opt portfolio				Smart-Opt portfolio			
Domestic country	α	MKT_{FX}	HML_{FX}	$\bar{R}^2$	α	MKT_{FX}	HML_{FX}	$\bar{R}^2$
Panel A: Full sample (1995:01 to 2019:03)								
United Kingdom	0.27%*	−0.25	0.73***	50.42%	0.69%***	−0.37	0.43**	16.56%
Switzerland	0.27%**	0.86***	0.65***	60.27%	0.47%**	0.95**	0.34*	24.69%
Japan	0.24%	0.42***	0.74***	54.92%	0.59%*	0.20	0.59***	16.76%
Canada	0.24%*	0.27***	0.80***	51.91%	0.63%***	0.48***	0.47***	16.89%
Australia	0.39%***	−0.36***	0.64***	51.05%	0.65%***	−0.40***	0.30	18.83%
New Zealand	0.15%	−0.66***	0.55***	56.20%	0.62%**	−0.99***	0.11	35.76%
Sweden	0.33%**	0.34***	0.75***	52.49%	0.65%***	0.59***	0.46***	20.22%
Norway	0.37%**	−0.36***	0.70***	45.13%	0.71%***	0.48***	0.45***	18.18%
Denmark	0.27%	−0.19	0.76***	42.93%	0.45%*	0.25	0.42**	12.84%
Panel B: Pre-crisis subsample (1995:01 to 2008:08)								
United Kingdom	0.32%**	−0.84***	0.87***	74.45%	0.48%**	−1.25***	0.79***	69.42%
Switzerland	0.33%*	0.83***	0.90***	67.44%	0.09%	1.17***	0.81***	68.23%
Japan	0.28%*	0.82***	0.79***	76.19%	0.14%	0.92***	0.83***	69.50%
Canada	0.30%**	0.40***	1.08***	64.81%	0.30%	0.41***	1.03***	48.97%
Australia	0.31%*	−0.08	1.05***	63.73%	0.31%*	−0.15	1.07***	57.66%
New Zealand	0.19%	−0.81***	0.75***	65.54%	0.15%	−1.04***	0.82***	59.62%
Sweden	0.44%***	−0.01	0.94***	65.50%	0.43%**	0.51***	0.98***	53.45%
Norway	0.51%**	−0.66***	0.86***	46.17%	0.71%***	0.21	0.83***	39.25%
Denmark	0.33%***	−0.41**	1.01***	67.53%	0.26%*	0.39	0.95***	62.65%
Panel C: Post-crisis subsample (2008:09 to 2019:03)								
United Kingdom	−0.04%	0.25*	0.53***	47.71%	0.48%**	0.41**	0.01	8.21%
Switzerland	0.02%	0.74***	0.46***	54.11%	0.54%*	0.52	−0.03	4.03%
Japan	−0.13%	0.02	0.74***	54.42%	0.55%*	−0.47*	0.47***	19.60%
Canada	−0.00%	0.19	0.58***	42.86%	0.62%**	0.90***	0.05	22.82%
Australia	0.17%	−0.80***	0.20	53.43%	0.49%**	−0.70***	−0.39	14.08%
New Zealand	−0.07%	−0.54***	0.36***	49.82%	0.64%*	−0.99***	−0.48*	36.12%
Sweden	0.01%	0.66***	0.61***	50.66%	0.60%**	0.52***	0.02	7.28%
Norway	−0.03%	−0.13	0.61***	50.40%	0.38%	0.55*	0.14	8.03%
Denmark	0.04%	−0.15	0.57***	26.36%	0.46%	−0.28	−0.04	−0.39%

The table reports Lustig, Roussanov, and Verdelhan (2011) currency factor model estimation results for the portfolio excess return for a non-US domestic investor with mean-variance preferences and a relative risk aversion coefficient of five who invests in 14 foreign currencies. The Basic-Opt (Smart-Opt) portfolio assumes that the investor uses the no-change (elastic net) exchange rate forecast. The elastic net forecast is based on elastic net estimation of a panel predictive regression with 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables, where the regularization parameter is tuned via the extended regularization information criterion. MKT_{FX} (HML_{FX}) is the dollar (carry trade) risk factor. Brackets report t-statistics based on robust standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table A11: Smart-Opt portfolio performance, DNN forecasts

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	DNN3 forecast			DNN4 forecast		
Transaction costs	Annualized mean	Annualized volatility	Annualized Sharpe ratio	Annualized mean	Annualized volatility	Annualized Sharpe ratio
Panel A: Full sample (1995:01 to 2019:03)						
Mid-quotes	12.51%	14.89%	0.84***	7.92%	12.86%	0.62***
25%	10.15%	14.99%	0.68***	5.73%	12.86%	0.45**
50%	9.27%	14.98%	0.62***	4.89%	12.85%	0.38*
75%	8.40%	14.97%	0.56***	4.05%	12.86%	0.32
Bid-ask spread	7.52%	14.96%	0.50**	3.21%	12.86%	0.25
Panel B: Pre-crisis subsample (1995:01 to 2008:08)						
Mid-quotes	14.12%	14.62%	0.97***	8.40%	13.33%	0.63**
25%	11.60%	14.80%	0.78***	6.01%	13.45%	0.45*
50%	10.49%	14.78%	0.71***	4.95%	13.45%	0.37
75%	9.37%	14.77%	0.63**	3.89%	13.45%	0.29
Bid-ask spread	8.26%	14.76%	0.56**	2.83%	13.45%	0.21
Panel C: Post-crisis subsample (2008:09 to 2019:03)						
Mid-quotes	10.43%	15.26%	0.68**	7.30%	12.27%	0.60*
25%	8.28%	15.27%	0.54*	5.38%	12.10%	0.44
50%	7.71%	15.27%	0.50	4.82%	12.10%	0.40
75%	7.14%	15.27%	0.47	4.27%	12.09%	0.35
Bid-ask spread	6.56%	15.27%	0.43	3.71%	12.09%	0.31

The table reports annualized summary statistics for the portfolio excess return for a US investor with mean-variance preferences and a relative risk aversion coefficient of five who invests in 14 foreign currencies. The Smart-Opt portfolio assumes that the investor uses the deep neural network (DNN) exchange rate forecast. The DNN forecast is based on 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables. DNN3 (DNN4) contains three (four) hidden layers. Mid-quotes ignore transaction costs. We account for transaction costs using bid-ask spreads from Datastream. Because the full bid-ask spreads likely overstate transaction costs, we compute results assuming 25%, 50%, and 75% of the full bid-ask spreads. We test the significance of the Sharpe ratios using t-statistics based on Bao (2009) standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table A12: Smart-Opt portfolio alphas and currency factor exposures, DNN forecasts

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	DNN3 forecast				DNN4 forecast			
Sample	Ann. alpha	MKT_{FX}	HML_{FX}	$\bar{R}^2$	Ann. alpha	MKT_{FX}	HML_{FX}	$\bar{R}^2$
1995:01 to 2019:03	12.05%	−0.22	0.11	0.80%	6.38%	0.03	0.31	5.41%
	[2.99]***	[−1.66]	[0.67]		[1.82]*	[0.18]	[2.11]**	
1995:01 to 2008:08	11.43%	−0.08	0.42	5.57%	3.39%	0.35	0.68	22.55%
	[2.37]**	[−0.28]	[2.42]**		[0.85]	[1.44]	[4.28]***	
2008:09 to 2019:03	10.75%	−0.15	−0.18	2.12%	7.12%	−0.06	0.05	−1.42%
	[1.80]*	[−0.60]	[−0.69]		[1.60]	[−0.32]	[0.26]	

The table reports Lustig, Roussanov, and Verdelhan (2011) currency factor model estimation results for the portfolio excess return for a US investor with mean-variance preferences and a relative risk aversion coefficient of five who invests in 14 foreign currencies. The Smart-Opt portfolio assumes that the investor uses the deep neural network (DNN) exchange rate forecast. The DNN forecast is based on 70 predictors formed from ten country characteristics and interactions of the ten country characteristics with six global variables. DNN3 (DNN4) contains three (four) hidden layers. Mid-quotes ignore transaction costs. MKT_{FX} (HML_{FX}) is the dollar (carry trade) risk factor. Brackets report t-statistics based on robust standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

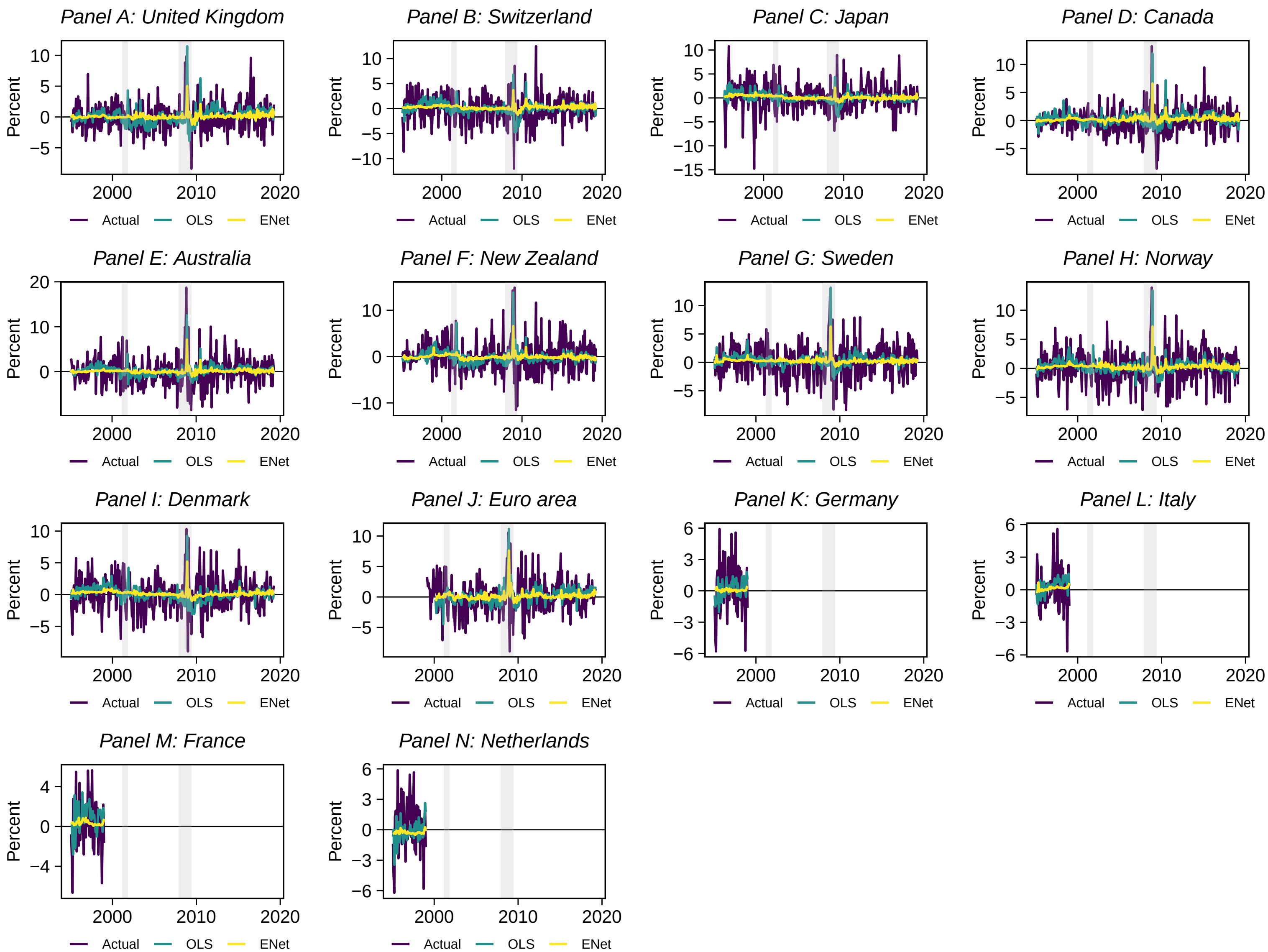


Figure A1: OLS and ENet-BIC forecasts

Each panel shows monthly out-of-sample forecasts of exchange rate changes based on recursive ordinary least squares (OLS) and elastic net (ENet) estimation of a panel predictive regression with 70 predictors, where the regularization parameter for the elastic net is tuned via the Bayesian information criterion (BIC). Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

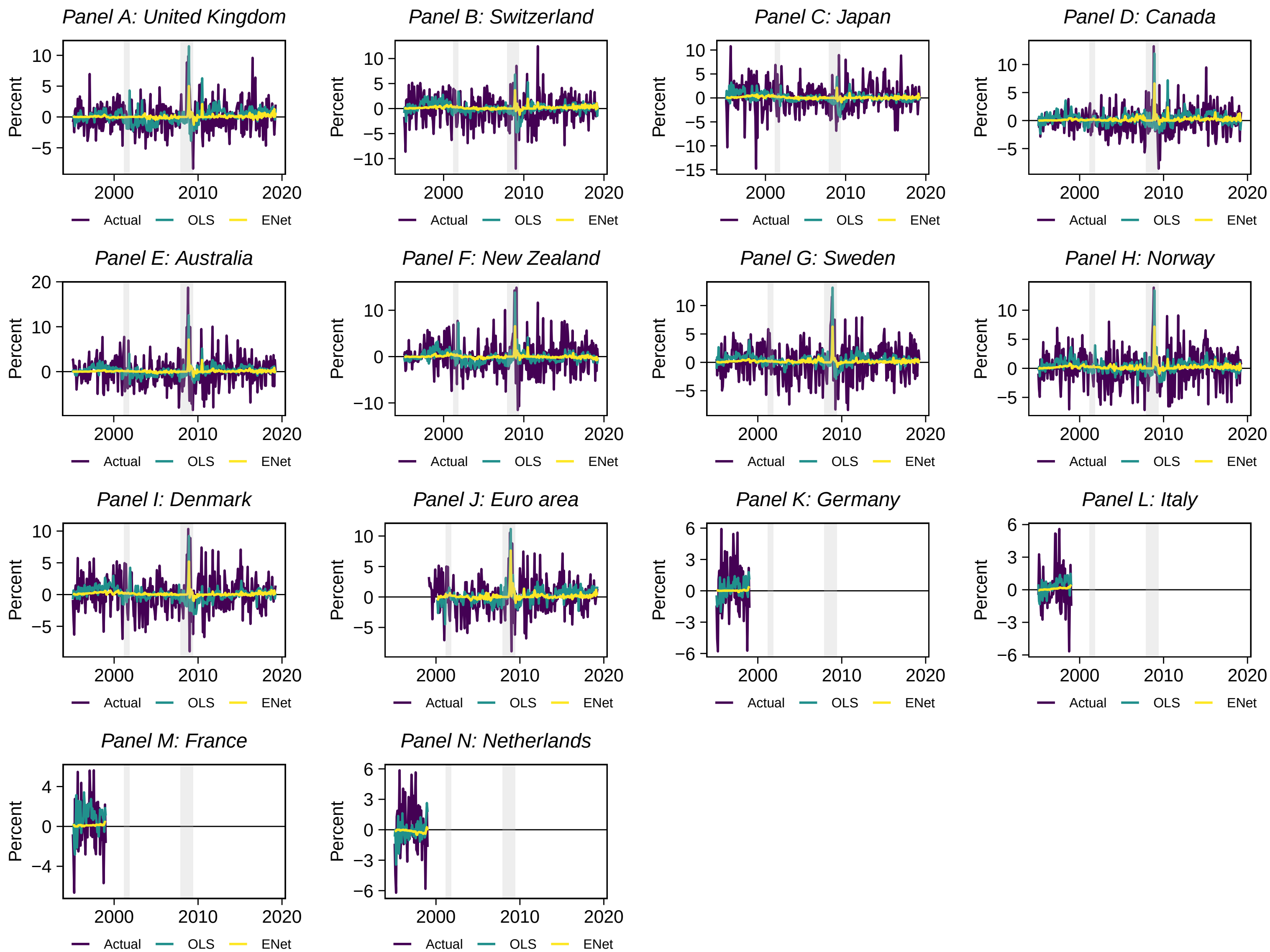


Figure A2: OLS and ENet-MBIC forecasts

Each panel shows monthly out-of-sample forecasts of exchange rate changes based on recursive ordinary least squares (OLS) and elastic net (ENet) estimation of a panel predictive regression with 70 predictors, where the regularization parameter for the elastic net is tuned via the modified Bayesian information criterion (MBIC). Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

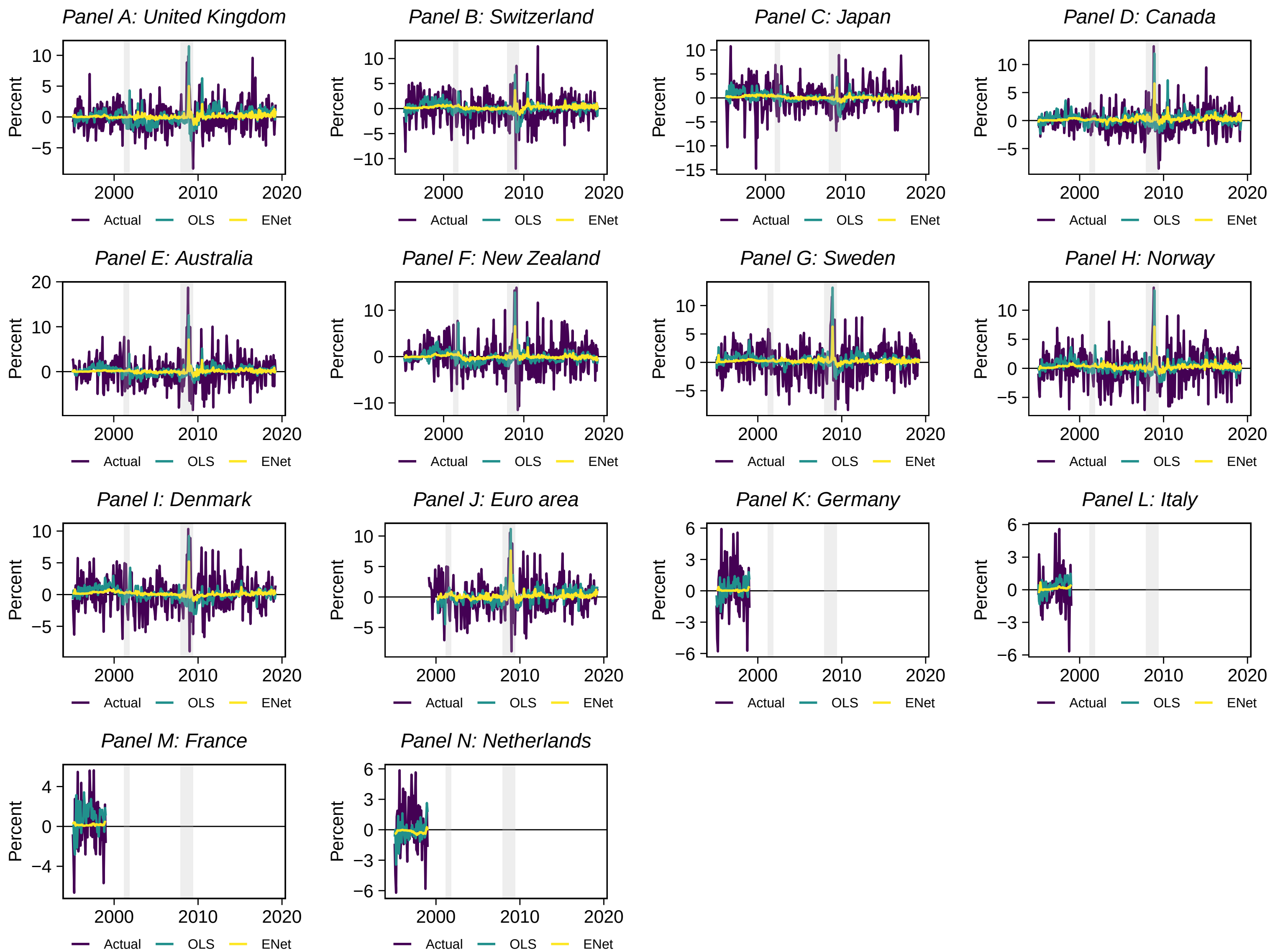


Figure A3: OLS and ENet-GIC forecasts

Each panel shows monthly out-of-sample forecasts of exchange rate changes based on recursive ordinary least squares (OLS) and elastic net (ENet) estimation of a panel predictive regression with 70 predictors, where the regularization parameter for the elastic net is tuned via the generalized information criterion (GIC). Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

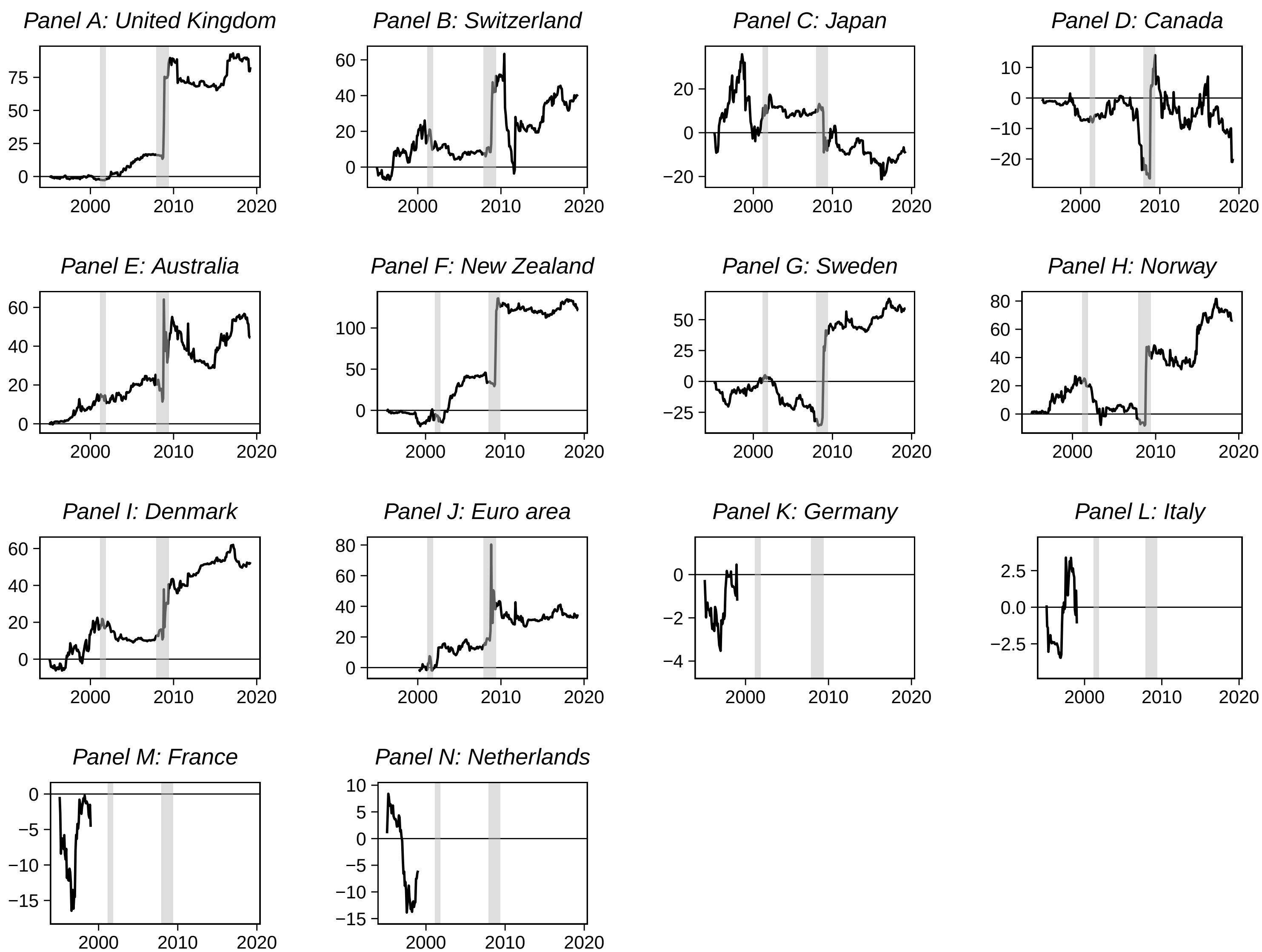


Figure A4: Cumulative difference in squared forecast errors for ENet-BIC forecast

The figure shows the cumulative difference in squared forecast errors for the no-change benchmark forecast vis-à-vis the ENet-BIC forecast. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

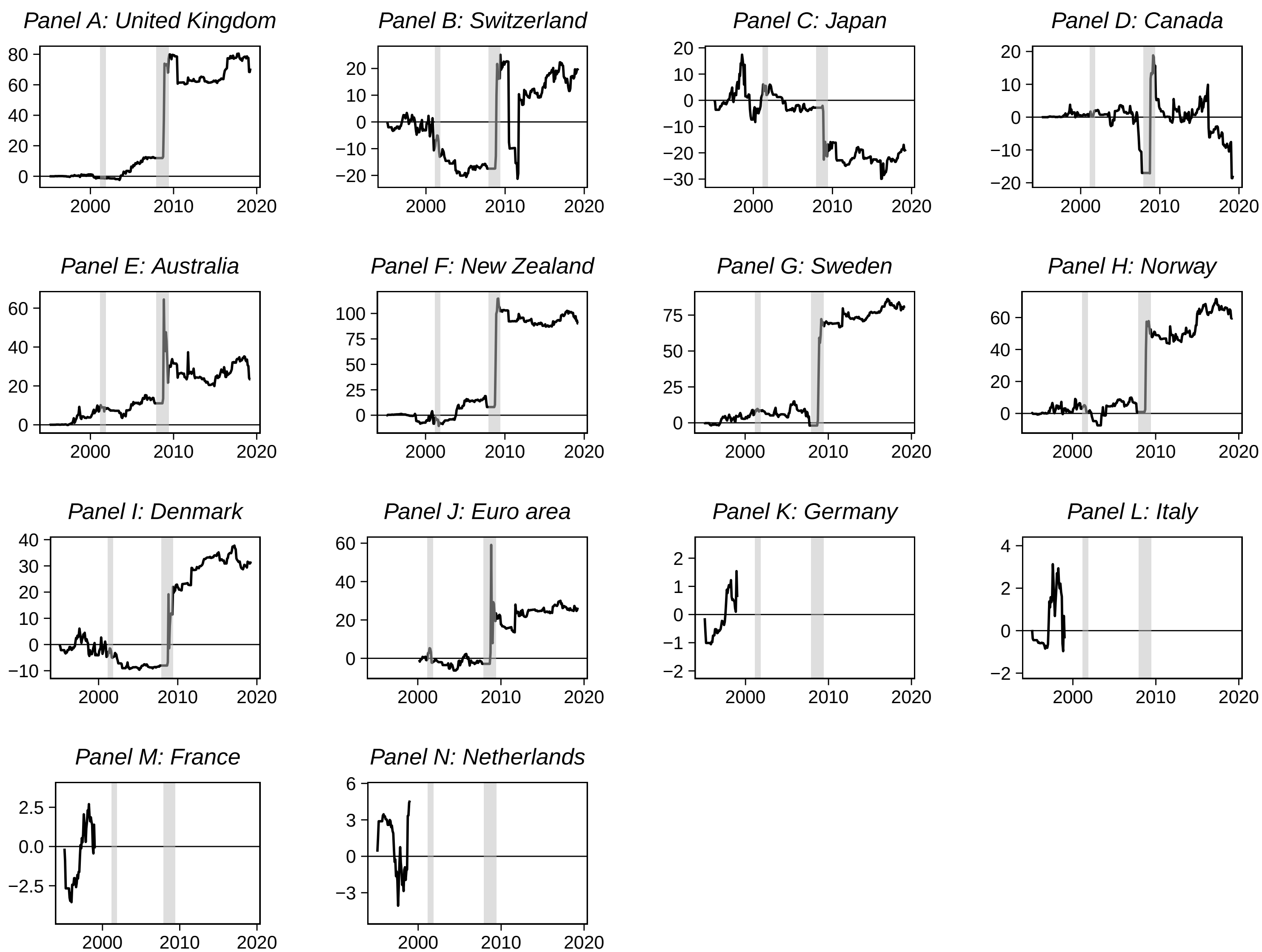


Figure A5: Cumulative difference in squared forecast errors for ENet-MBIC forecast

The figure shows the cumulative difference in squared forecast errors for the no-change benchmark forecast vis-à-vis the ENet-MBIC forecast. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

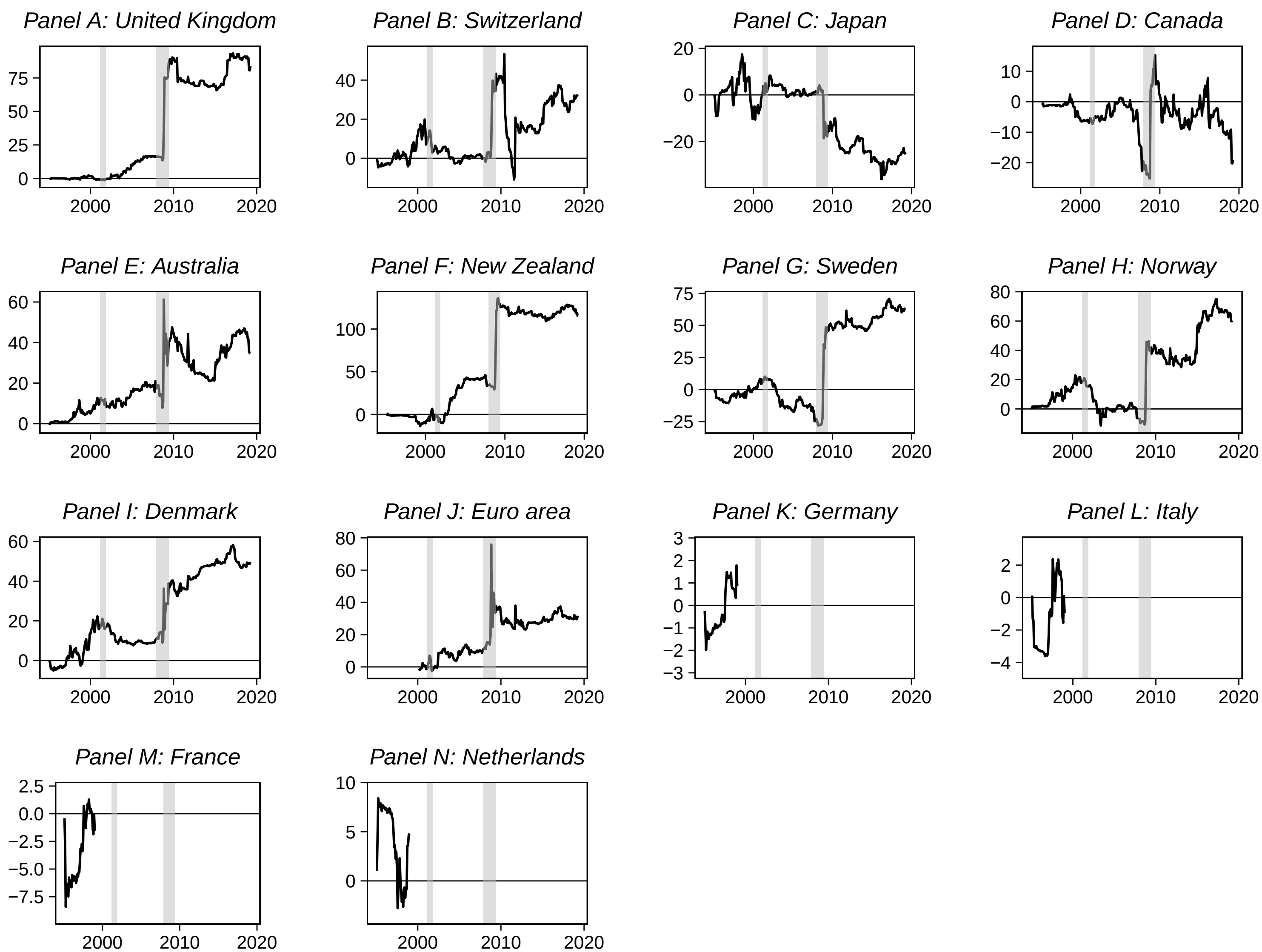


Figure A6: Cumulative difference in squared forecast errors for ENet-GIC forecast

The figure shows the cumulative difference in squared forecast errors for the no-change benchmark forecast vis-à-vis the ENet-GIC forecast. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

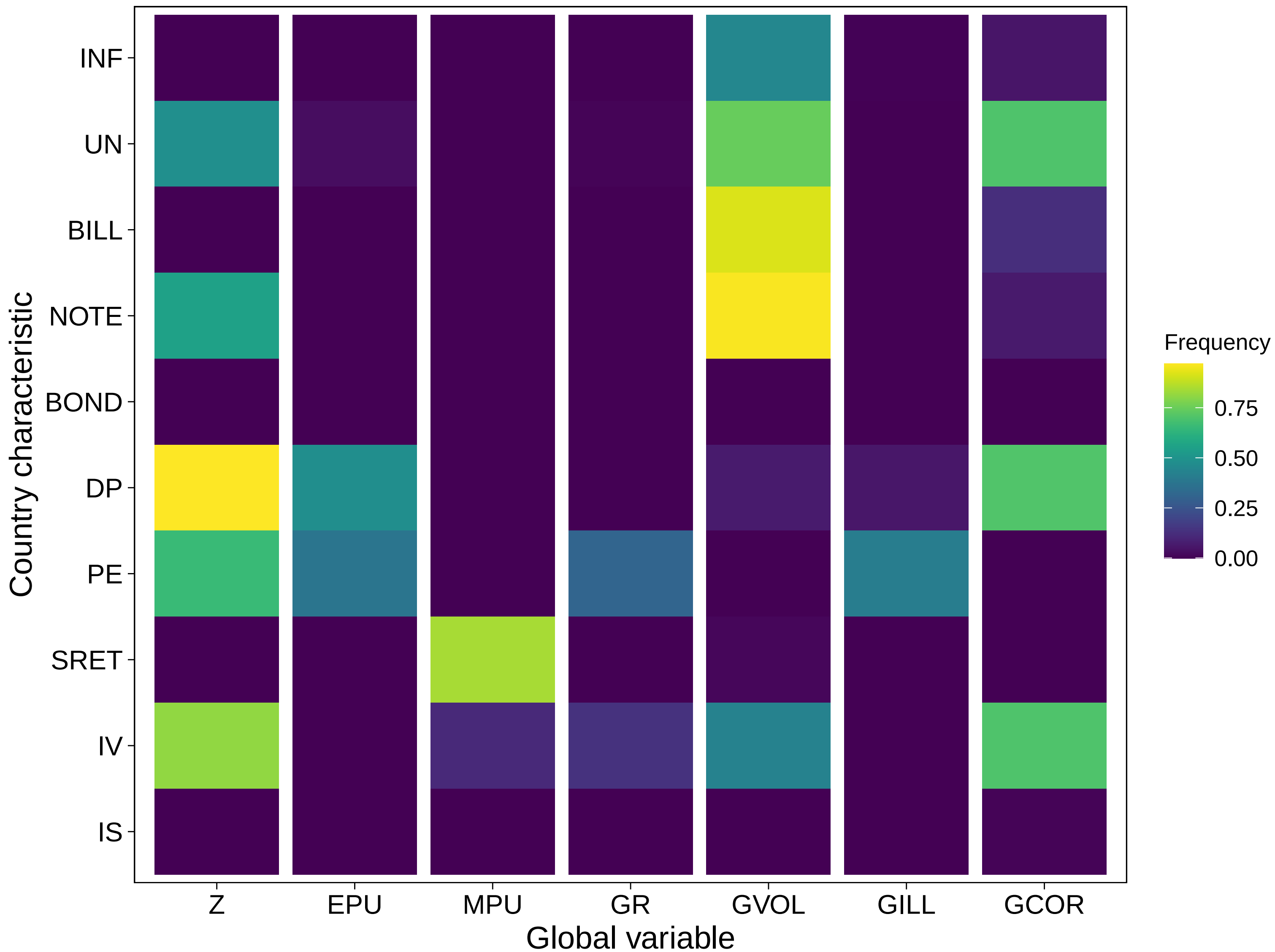


Figure A7: Heatmap for predictors selected by the elastic net via BIC tuning

The figure shows the frequencies for predictors selected by the elastic net when the regularization parameter is tuned via the Bayesian information criterion (BIC). The Z column gives the frequencies for the individual country characteristics; the other columns give the frequencies for country characteristics interacted with the global variables.

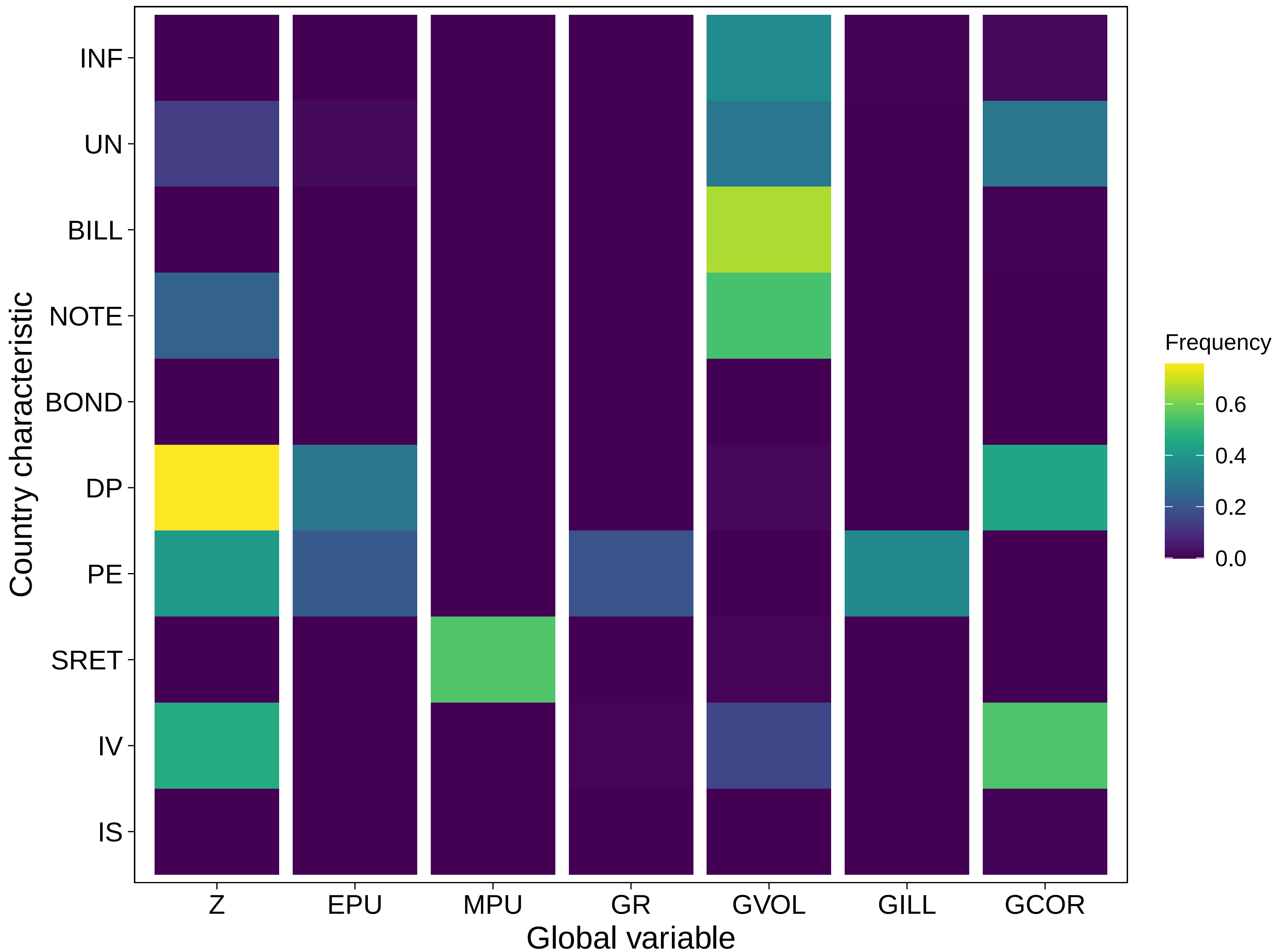


Figure A8: Heatmap for predictors selected by the elastic net via MBIC tuning

The figure shows the frequencies for predictors selected by the elastic net when the regularization parameter is tuned via the modified Bayesian information criterion (MBIC). The Z column gives the frequencies for the individual country characteristics; the other columns give the frequencies for country characteristics interacted with the global variables.

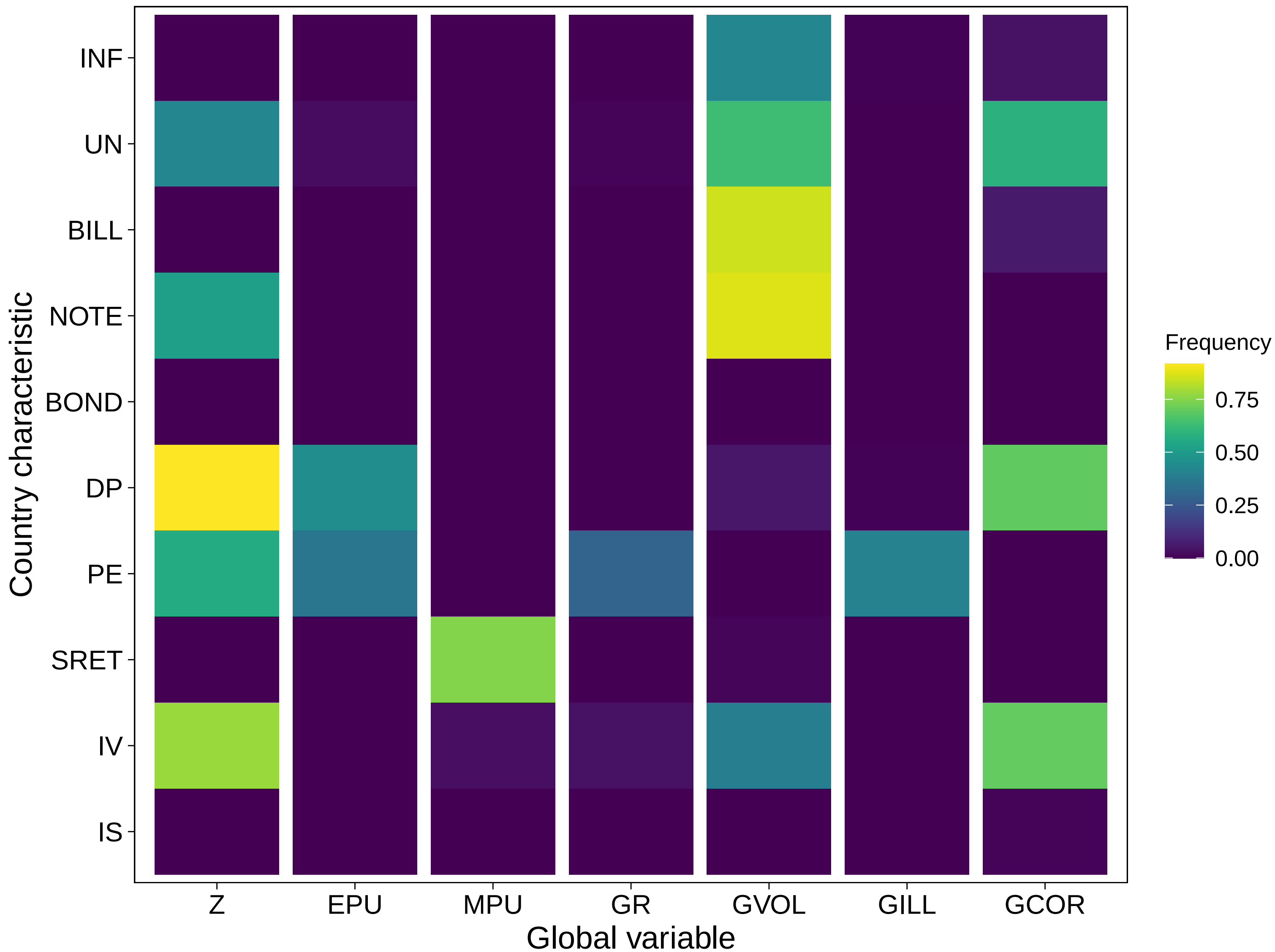


Figure A9: Heatmap for predictors selected by the elastic net via GIC tuning

The figure shows the frequencies for predictors selected by the elastic net when the regularization parameter is tuned via the generalized information criterion (GIC). The Z column gives the frequencies for the individual country characteristics; the other columns give the frequencies for country characteristics interacted with the global variables.

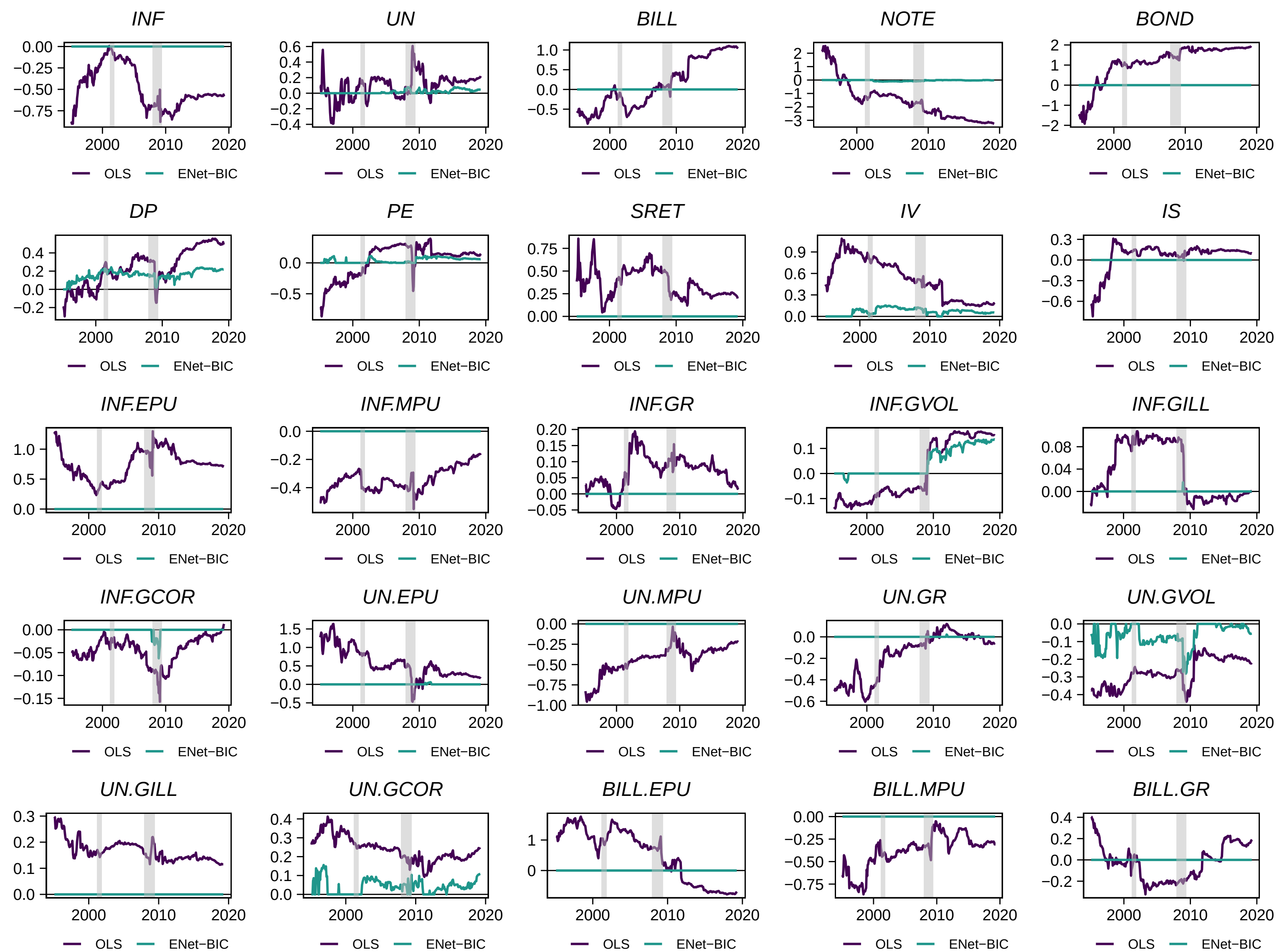


Figure A10: OLS and ENet-BIC recursive coefficient estimates

The figure shows recursive ordinary least squares (OLS) and elastic net (ENet) slope coefficient estimates for a panel predictive regression with 70 predictors, where the regularization parameter for the elastic net is tuned via the Bayesian information criterion (BIC). Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

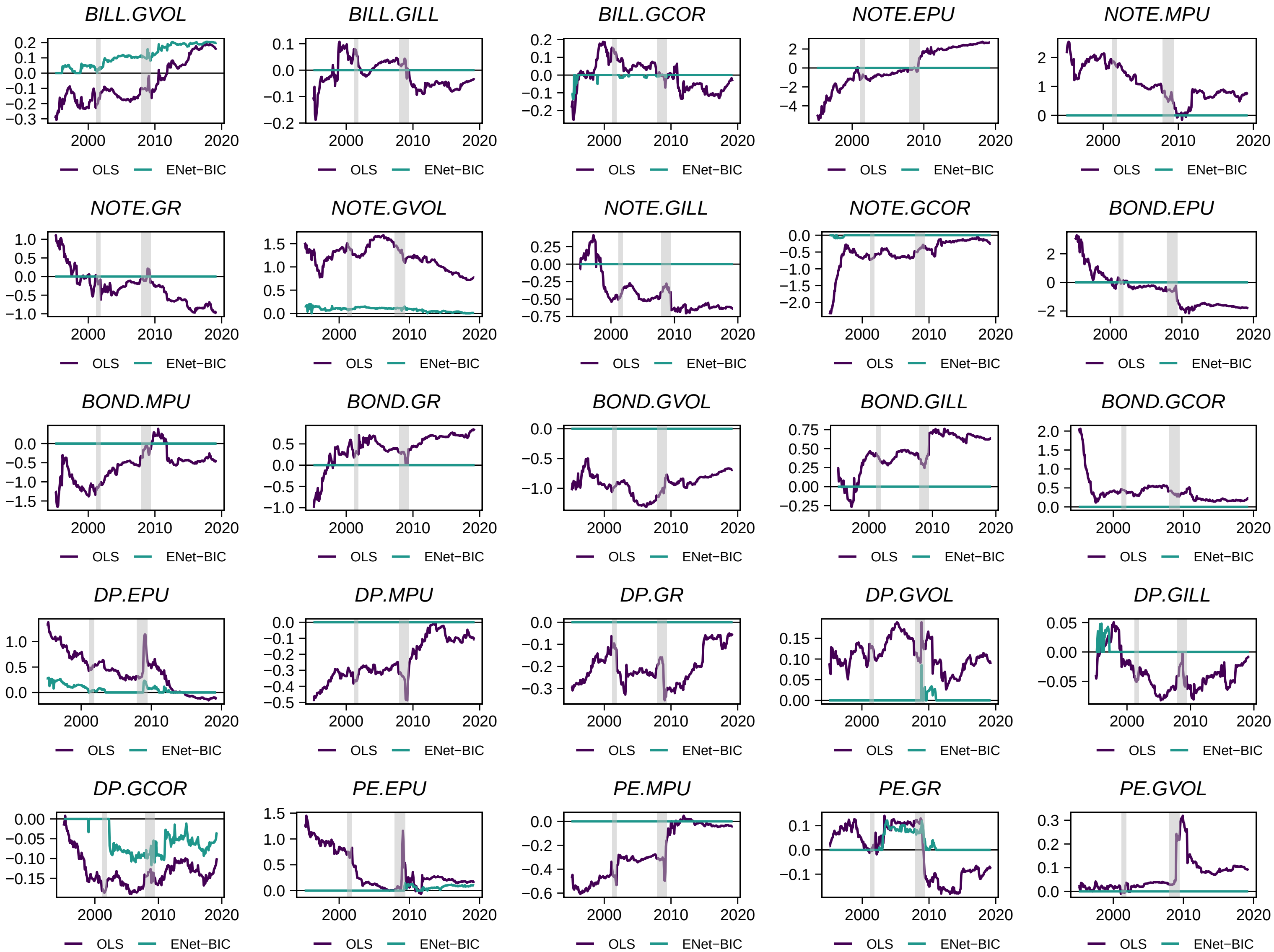


Figure A10 (continued)

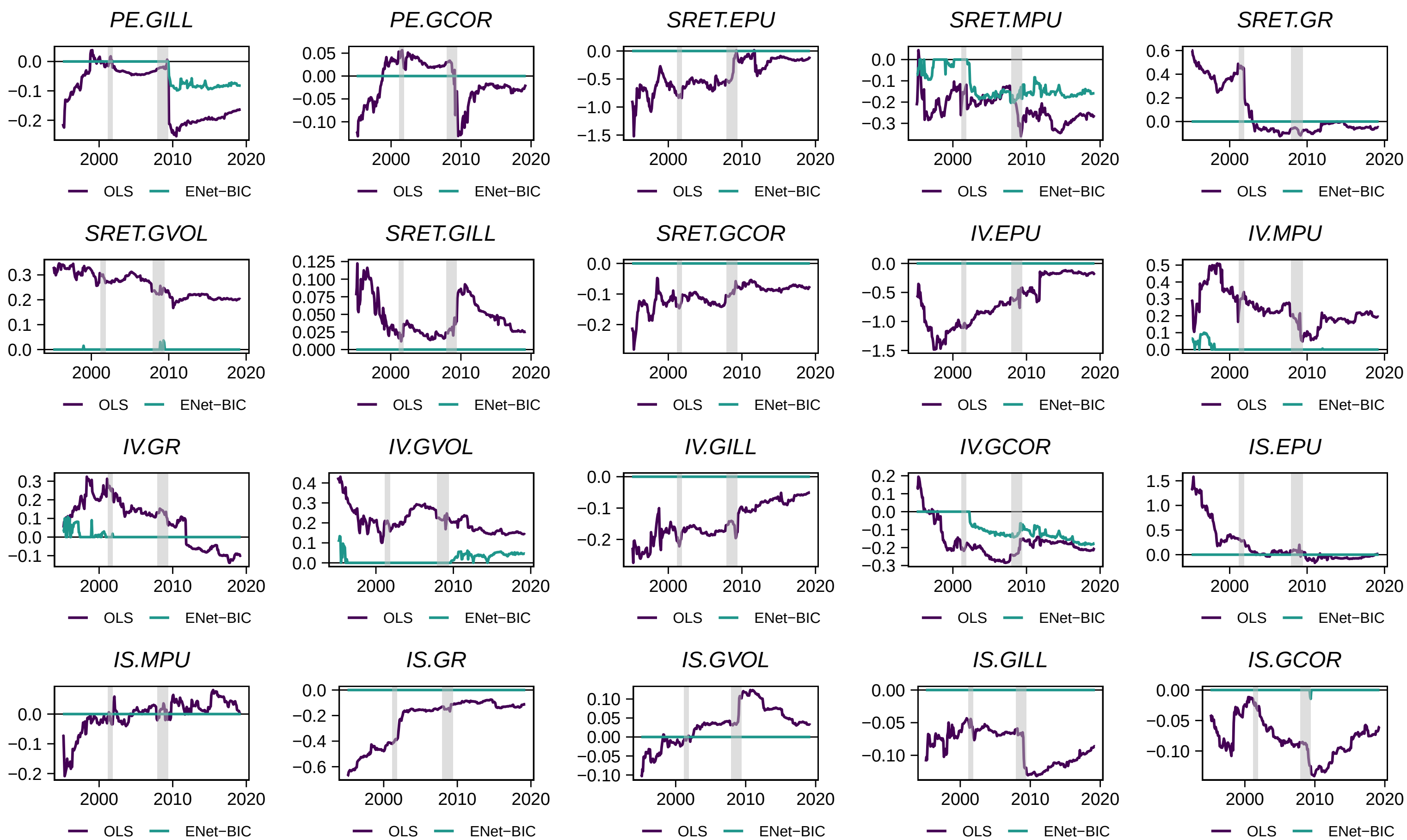


Figure A10 (continued)

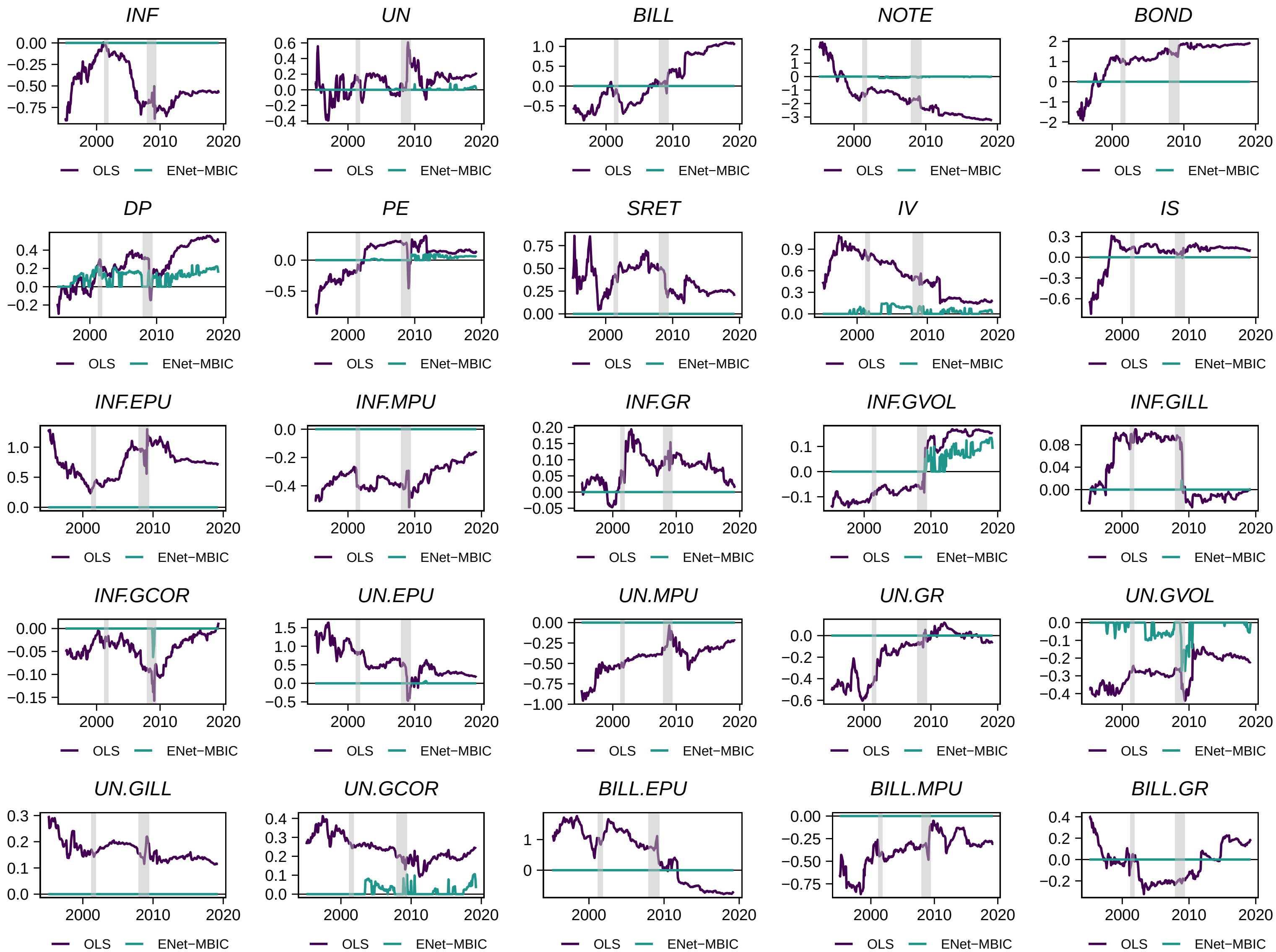


Figure A11: OLS and ENet-MBIC recursive coefficient estimates

The figure shows recursive ordinary least squares (OLS) and elastic net (ENet) slope coefficient estimates for a panel predictive regression with 70 predictors, where the regularization parameter for the elastic net is tuned via the modified Bayesian information criterion (MBIC). Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

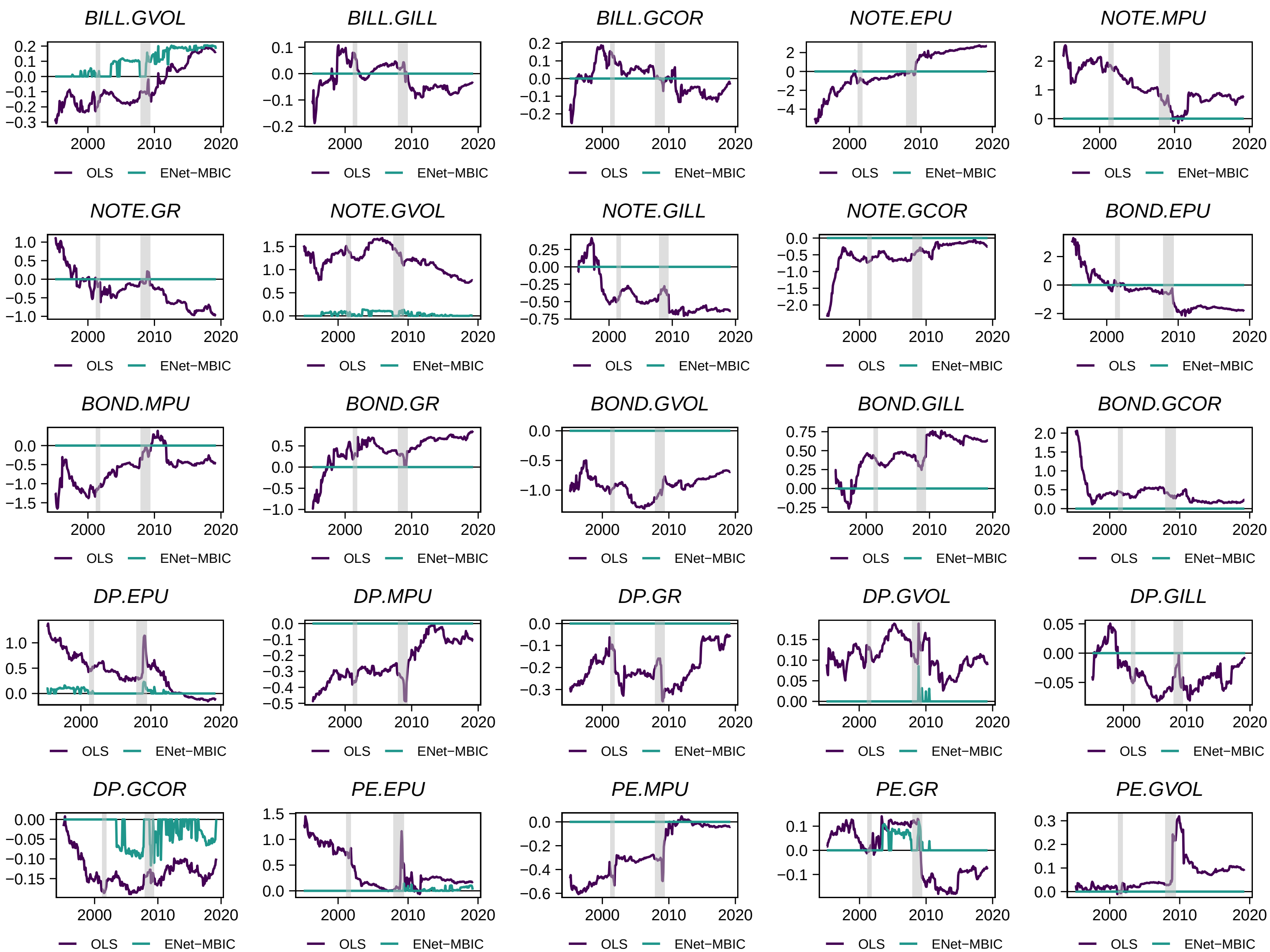


Figure A11 (continued)

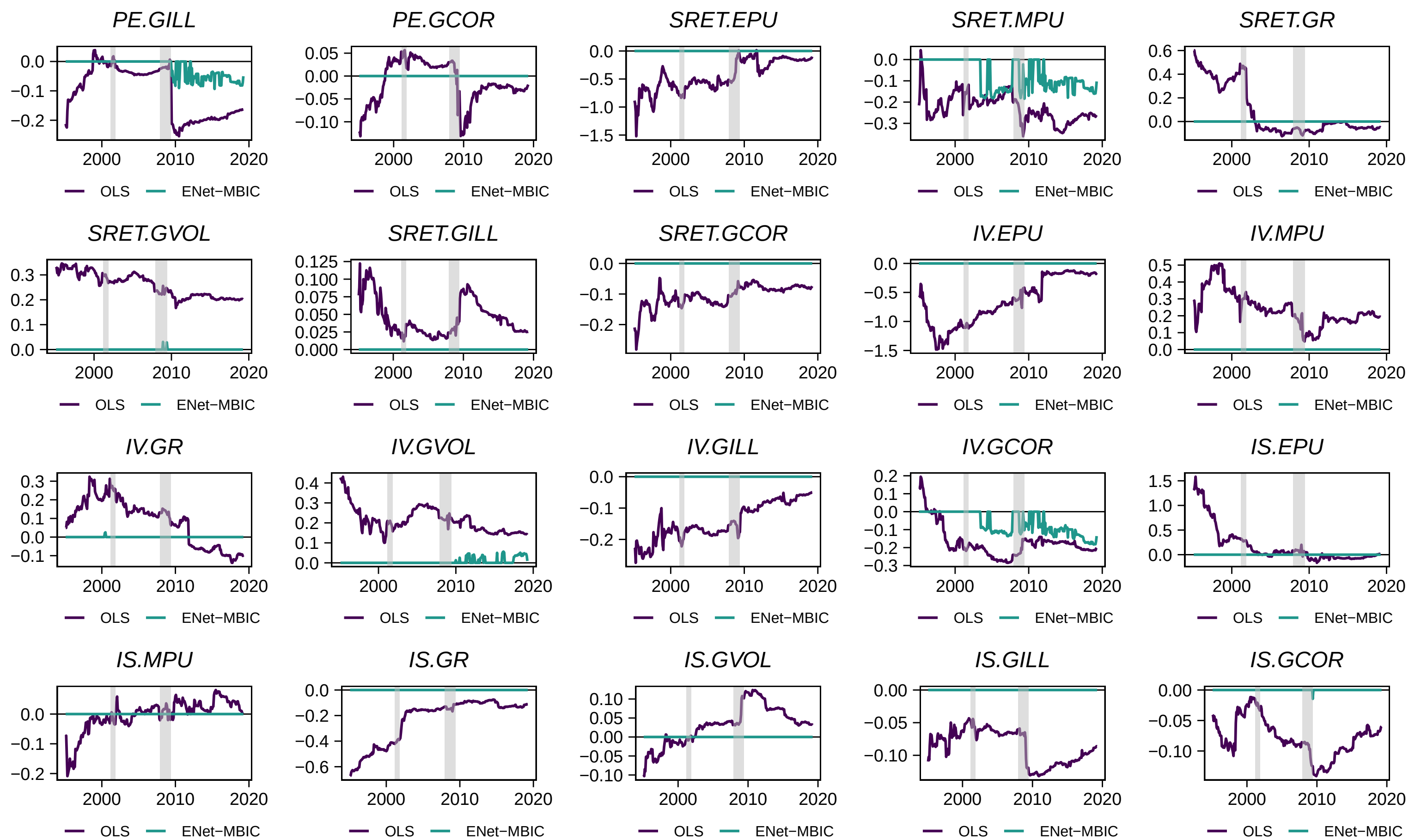


Figure A11 (continued)

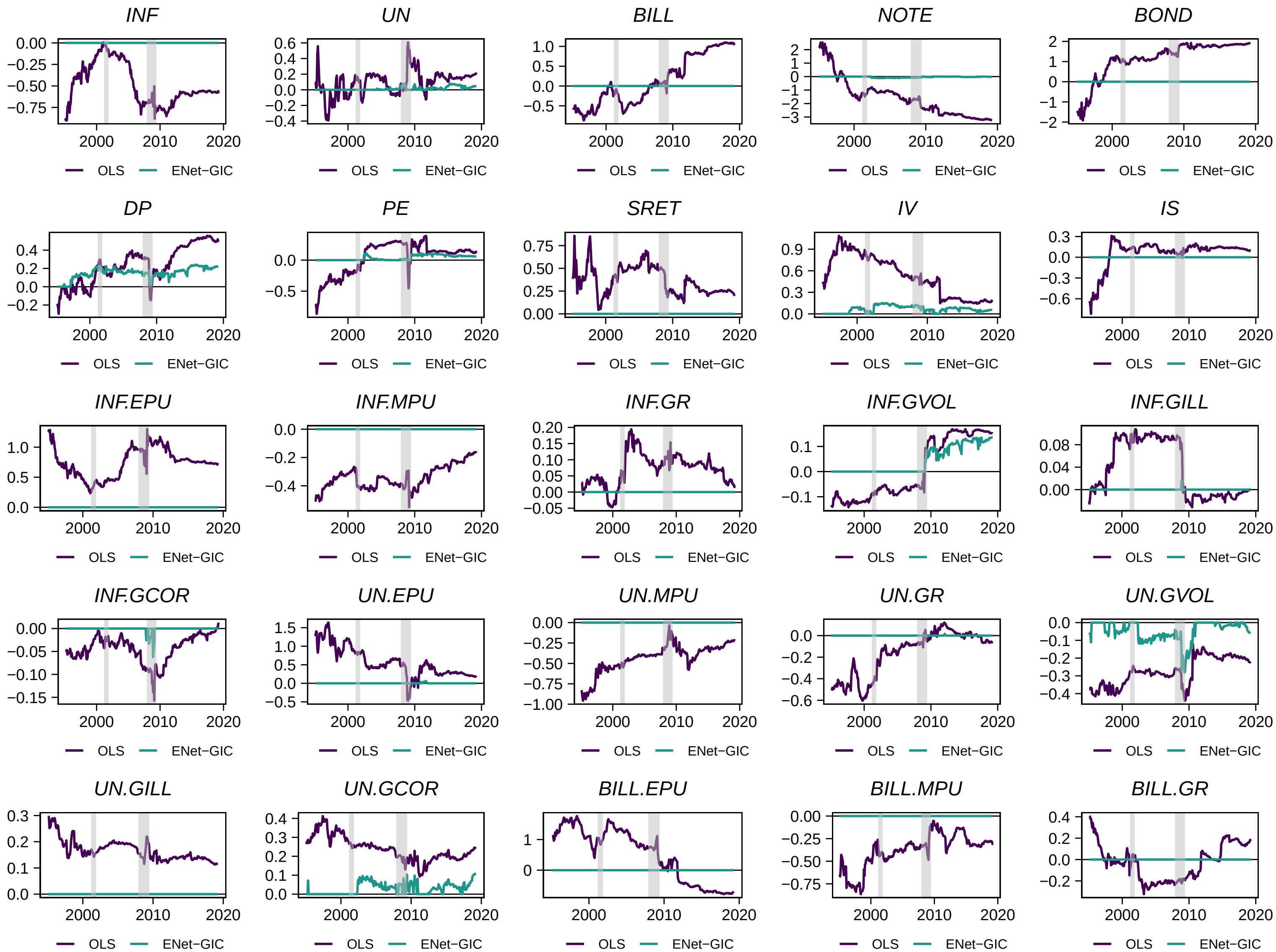


Figure A12: OLS and ENet-GIC recursive coefficient estimates

The figure shows recursive ordinary least squares (OLS) and elastic net (ENet) slope coefficient estimates for a panel predictive regression with 70 predictors, where the regularization parameter for the elastic net is tuned via the generalized information criterion (GIC). Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

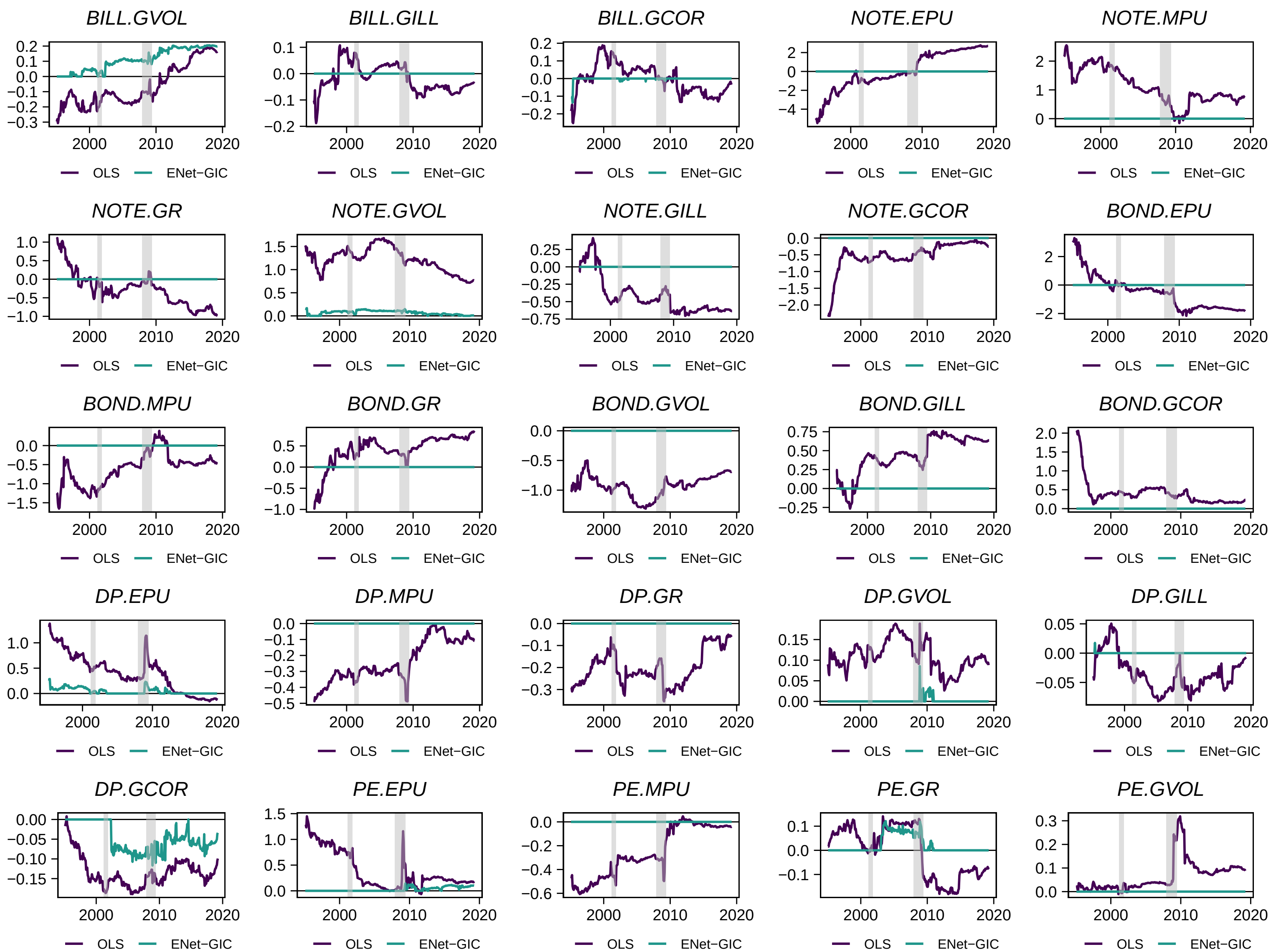


Figure A12 (continued)

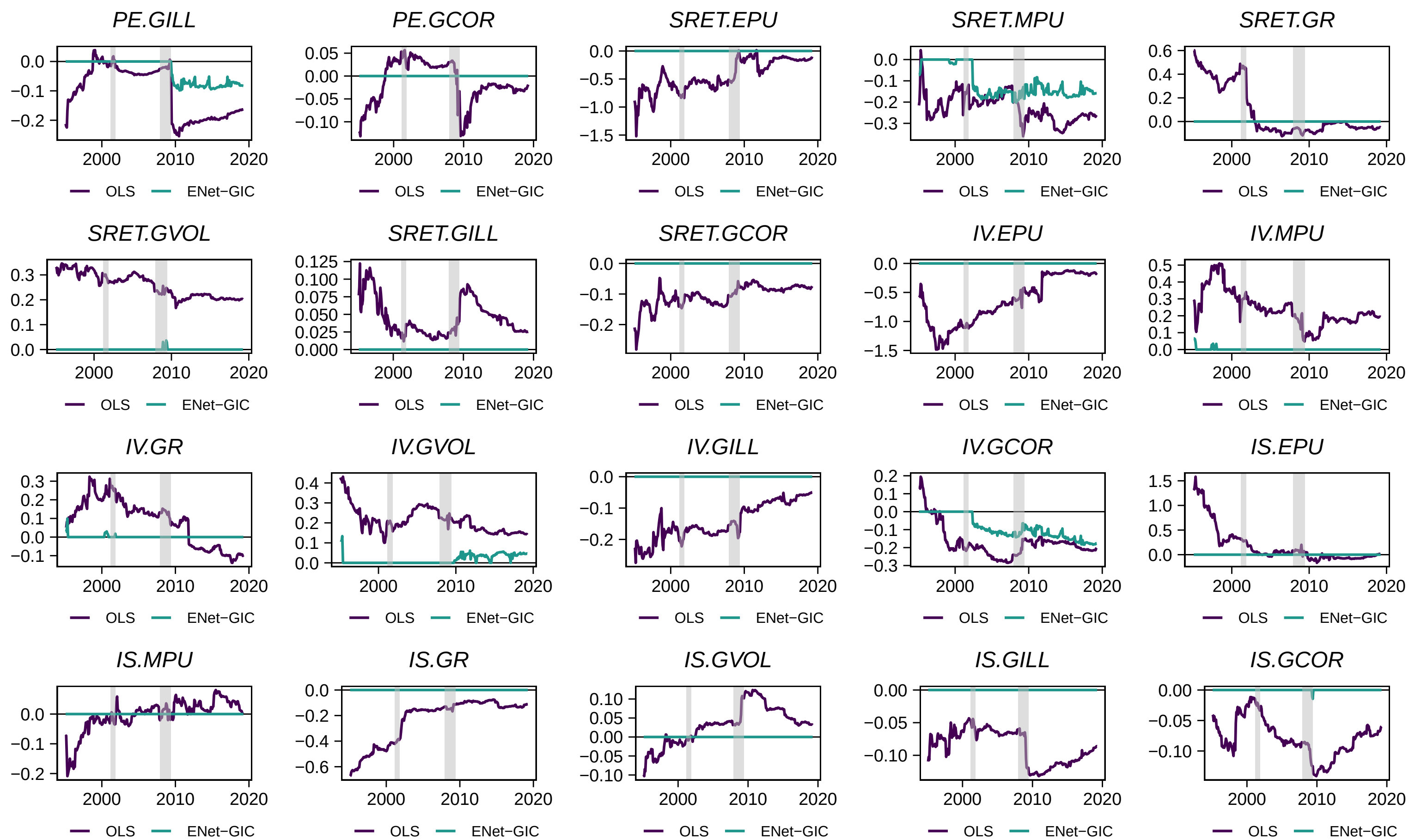


Figure A12 (continued)